

**eval**

The main task, comprising all test items.

number of items in task:	8448
--------------------------	------

	all senses	main senses only
average polysemy:	10.372	7.207

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.916	1.512

**All systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.965</b> (0.963)	<b>0.968</b> (0.967)	<b>0.970</b> (0.968)
best system	<b>0.771</b> (0.771)	<b>0.797</b> (0.797)	<b>0.814</b> (0.813)
average of systems	<b>0.550</b> (0.376)	<b>0.632</b> (0.410)	<b>0.661</b> (0.426)
worst system	<b>0.205</b> (0.162)	<b>0.315</b> (0.248)	<b>0.338</b> (0.267)
best <u>baseline</u>	<b>0.691</b> (0.689)	<b>0.720</b> (0.719)	<b>0.741</b> (0.739)

**A systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.559</b> (0.519)	<b>0.616</b> (0.031)	<b>0.650</b> (0.198)
average of systems	<b>0.406</b> (0.222)	<b>0.531</b> (0.264)	<b>0.563</b> (0.280)
worst system	<b>0.205</b> (0.162)	<b>0.315</b> (0.248)	<b>0.338</b> (0.267)

best <u>baseline</u>	<b>0.550</b> (0.548)	<b>0.584</b> (0.582)	<b>0.600</b> (0.597)
----------------------	----------------------	----------------------	----------------------

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.771</b> (0.771)	<b>0.797</b> (0.797)	<b>0.814</b> (0.813)
average of systems	<b>0.661</b> (0.540)	<b>0.707</b> (0.570)	<b>0.733</b> (0.588)
worst system	<b>0.433</b> (0.138)	<b>0.588</b> (0.188)	<b>0.627</b> (0.200)
best <u>baseline</u>	<b>0.691</b> (0.689)	<b>0.720</b> (0.719)	<b>0.741</b> (0.739)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.649</b> (0.127)	<b>0.733</b> (0.143)	<b>0.767</b> (0.150)
average of systems	<b>0.527</b> (0.091)	<b>0.637</b> (0.109)	<b>0.663</b> (0.113)
worst system	<b>0.405</b> (0.056)	<b>0.540</b> (0.075)	<b>0.559</b> (0.077)

Detailed results**trainable**

The subset of test items in files of words for which corpus training data was supplied.

number of items in task:	<b>7446</b>
--------------------------	-------------

	all senses	main senses only
average polysemy:	10.788	7.432

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
average <u><a href="#">entropy</a></u> :	1.962	1.551

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.963</b> (0.962)	<b>0.967</b> (0.966)	<b>0.969</b> (0.967)
best system	<b>0.763</b> (0.756)	<b>0.788</b> (0.780)	<b>0.801</b> (0.793)
average of systems	<b>0.548</b> (0.393)	<b>0.622</b> (0.424)	<b>0.655</b> (0.442)
worst system	<b>0.199</b> (0.156)	<b>0.306</b> (0.240)	<b>0.333</b> (0.261)
best <u><a href="#">baseline</a></u>	<b>0.709</b> (0.708)	<b>0.735</b> (0.734)	<b>0.759</b> (0.758)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.574</b> (0.533)	<b>0.617</b> (0.572)	<b>0.648</b> (0.600)
average of systems	<b>0.400</b> (0.219)	<b>0.512</b> (0.256)	<b>0.551</b> (0.275)
worst system	<b>0.199</b> (0.156)	<b>0.306</b> (0.240)	<b>0.333</b> (0.261)
best <u><a href="#">baseline</a></u>	<b>0.549</b> (0.547)	<b>0.582</b> (0.579)	<b>0.599</b> (0.596)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.763</b> (0.756)	<b>0.788</b> (0.780)	<b>0.801</b> (0.793)
average of systems	<b>0.666</b> (0.574)	<b>0.705</b> (0.601)	<b>0.733</b> (0.623)
worst system	<b>0.438</b> (0.128)	<b>0.564</b> (0.165)	<b>0.612</b> (0.179)
best <u><a href="#">baseline</a></u>	<b>0.709</b> (0.708)	<b>0.735</b> (0.734)	<b>0.759</b> (0.758)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.649</b> (0.144)	<b>0.733</b> (0.162)	<b>0.767</b> (0.170)
average of systems	<b>0.506</b> (0.093)	<b>0.622</b> (0.111)	<b>0.651</b> (0.117)
worst system	<b>0.363</b> (0.043)	<b>0.510</b> (0.060)	<b>0.535</b> (0.063)

Detailed results**untrainable**

The complement of trainable.

number of items in task:	1002
--------------------------	------

	all senses	main senses only
average polysemy:	7.276	5.533

	fine-grained	coarse-grained
average <u>entropy</u> :	1.568	1.220

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.973</b> (0.973)	<b>0.974</b> (0.974)	<b>0.975</b> (0.975)
best system	<b>0.882</b> (0.882)	<b>0.930</b> (0.930)	<b>0.930</b> (0.930)
average of systems	<b>0.512</b> (0.343)	<b>0.660</b> (0.414)	<b>0.660</b> (0.415)

worst system	<b>0.248</b> (0.207)	<b>0.375</b> (0.312)	<b>0.376</b> (0.313)
best <u>baseline</u>	<b>0.626</b> (0.626)	<b>0.709</b> (0.709)	<b>0.711</b> (0.711)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.559</b> (0.228)	<b>0.742</b> (0.066)	<b>0.742</b> (0.066)
average of systems	<b>0.435</b> (0.243)	<b>0.618</b> (0.320)	<b>0.618</b> (0.321)
worst system	<b>0.248</b> (0.207)	<b>0.375</b> (0.312)	<b>0.376</b> (0.313)
best <u>baseline</u>	<b>0.626</b> (0.626)	<b>0.709</b> (0.709)	<b>0.711</b> (0.711)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.882</b> (0.882)	<b>0.930</b> (0.930)	<b>0.930</b> (0.930)
average of systems	<b>0.608</b> (0.498)	<b>0.717</b> (0.569)	<b>0.719</b> (0.570)
worst system	<b>0.415</b> (0.216)	<b>0.590</b> (0.587)	<b>0.592</b> (0.589)
best <u>baseline</u>	<b>0.556</b> (0.553)	<b>0.604</b> (0.602)	<b>0.606</b> (0.603)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.533</b> (0.154)	<b>0.631</b> (0.183)	<b>0.631</b> (0.183)
average of systems	<b>0.533</b> (0.154)	<b>0.631</b> (0.183)	<b>0.631</b> (0.183)
worst system	<b>0.533</b> (0.154)	<b>0.631</b> (0.183)	<b>0.631</b> (0.183)

Detailed results

**multi-word**

The subset of test items tagged with sense tags for word forms that are not derivable from the root form of a test word by a regular morphological process.

number of items in task:	800
--------------------------	-----

	all senses	main senses only
average polysemy:	16.862	12.594

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.499	2.013

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.984</b> (0.982)	<b>0.986</b> (0.984)	<b>0.986</b> (0.984)
best system	<b>0.897</b> (0.870)	<b>0.914</b> (0.886)	<b>0.921</b> (0.894)
average of systems	<b>0.529</b> (0.399)	<b>0.592</b> (0.432)	<b>0.671</b> (0.471)
worst system	<b>0.000</b> (0.000)	<b>0.200</b> (0.002)	<b>0.347</b> (0.291)
best <u>baseline</u>	<b>0.815</b> (0.658)	<b>0.832</b> (0.672)	<b>0.861</b> (0.696)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.572</b> (0.309)	<b>0.689</b> (0.372)	<b>0.794</b> (0.429)
average of systems	<b>0.289</b> (0.197)	<b>0.381</b> (0.229)	<b>0.515</b> (0.279)

worst system	<b>0.000</b> (0.000)	<b>0.200</b> (0.002)	<b>0.347</b> (0.291)
best <u>baseline</u>	<b>0.801</b> (0.800)	<b>0.823</b> (0.822)	<b>0.843</b> (0.842)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.897</b> (0.870)	<b>0.914</b> (0.886)	<b>0.921</b> (0.894)
average of systems	<b>0.675</b> (0.583)	<b>0.722</b> (0.620)	<b>0.765</b> (0.655)
worst system	<b>0.296</b> (0.141)	<b>0.337</b> (0.160)	<b>0.370</b> (0.176)
best <u>baseline</u>	<b>0.815</b> (0.658)	<b>0.832</b> (0.672)	<b>0.861</b> (0.696)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.739</b> (0.238)	<b>0.783</b> (0.252)	<b>0.831</b> (0.268)
average of systems	<b>0.735</b> (0.204)	<b>0.765</b> (0.213)	<b>0.808</b> (0.226)
worst system	<b>0.731</b> (0.170)	<b>0.747</b> (0.174)	<b>0.785</b> (0.183)

Detailed results**simple-word**

The complement of multi-word.

number of items in task:	7648
--------------------------	------

	all senses	main senses only
average polysemy:	9.693	6.644

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
average entropy:	1.855	1.460

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.963</b> (0.961)	<b>0.966</b> (0.965)	<b>0.968</b> (0.966)
best system	<b>0.761</b> (0.761)	<b>0.787</b> (0.787)	<b>0.804</b> (0.804)
average of systems	<b>0.548</b> (0.374)	<b>0.634</b> (0.408)	<b>0.660</b> (0.422)
worst system	<b>0.200</b> (0.156)	<b>0.320</b> (0.251)	<b>0.338</b> (0.265)
best <u><a href="#">baseline</a></u>	<b>0.678</b> (0.677)	<b>0.709</b> (0.708)	<b>0.730</b> (0.729)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.576</b> (0.544)	<b>0.630</b> (0.595)	<b>0.677</b> (0.194)
average of systems	<b>0.411</b> (0.224)	<b>0.541</b> (0.267)	<b>0.569</b> (0.280)
worst system	<b>0.200</b> (0.156)	<b>0.320</b> (0.251)	<b>0.338</b> (0.265)
best <u><a href="#">baseline</a></u>	<b>0.524</b> (0.522)	<b>0.564</b> (0.564)	<b>0.605</b> (0.605)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.761</b> (0.761)	<b>0.787</b> (0.787)	<b>0.804</b> (0.804)
average of systems	<b>0.662</b> (0.536)	<b>0.708</b> (0.564)	<b>0.732</b> (0.581)
worst system	<b>0.456</b> (0.138)	<b>0.611</b> (0.538)	<b>0.623</b> (0.548)



best <u>baseline</u>	<b>0.678</b> (0.677)	<b>0.709</b> (0.708)	<b>0.730</b> (0.729)
----------------------	----------------------	----------------------	----------------------

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.633</b> (0.115)	<b>0.724</b> (0.132)	<b>0.756</b> (0.137)
average of systems	<b>0.488</b> (0.080)	<b>0.612</b> (0.098)	<b>0.636</b> (0.101)
worst system	<b>0.344</b> (0.044)	<b>0.501</b> (0.064)	<b>0.516</b> (0.066)

Detailed results**unassignable**

The subset of test items tagged with an UNASSIGNABLE sense tag in the key.

number of items in task:	35
--------------------------	----

	all senses	main senses only
average polysemy:	11.771	8.086

	fine-grained	coarse-grained
average <u>entropy</u> :	2.415	1.946

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.938</b> (0.857)	<b>0.938</b> (0.857)	<b>0.938</b> (0.857)
best system	<b>0.400</b> (0.400)	<b>0.509</b> (0.087)	<b>0.509</b> (0.087)

average of systems	<b>0.261</b> (0.166)	<b>0.317</b> (0.190)	<b>0.324</b> (0.196)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.343</b> (0.343)	<b>0.371</b> (0.371)	<b>0.371</b> (0.371)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.364</b> (0.138)	<b>0.509</b> (0.087)	<b>0.509</b> (0.087)
average of systems	<b>0.230</b> (0.126)	<b>0.296</b> (0.151)	<b>0.299</b> (0.153)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.343</b> (0.343)	<b>0.371</b> (0.371)	<b>0.371</b> (0.371)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.400</b> (0.400)	<b>0.500</b> (0.086)	<b>0.500</b> (0.086)
average of systems	<b>0.311</b> (0.222)	<b>0.368</b> (0.251)	<b>0.379</b> (0.260)
worst system	<b>0.185</b> (0.143)	<b>0.232</b> (0.186)	<b>0.250</b> (0.200)
best <u>baseline</u>	<b>0.200</b> (0.200)	<b>0.250</b> (0.186)	<b>0.269</b> (0.200)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.205</b> (0.010)	<b>0.205</b> (0.010)	<b>0.205</b> (0.010)
average of systems	<b>0.102</b> (0.005)	<b>0.102</b> (0.005)	<b>0.102</b> (0.005)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)

Detailed results**proper**

The subset of test items tagged with a PROPER NOUN sense tag in the key.

number of items in task:	286
--------------------------	-----

	all senses	main senses only
average polysemy:	12.014	8.899

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.035	1.671

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.991</b> (0.981)	<b>0.993</b> (0.983)	<b>0.993</b> (0.983)
best system	<b>0.884</b> (0.881)	<b>0.930</b> (0.927)	<b>0.930</b> (0.927)
average of systems	<b>0.510</b> (0.267)	<b>0.606</b> (0.305)	<b>0.613</b> (0.308)
worst system	<b>0.026</b> (0.002)	<b>0.263</b> (0.017)	<b>0.263</b> (0.017)
best <u>baseline</u>	<b>0.756</b> (0.325)	<b>0.756</b> (0.325)	<b>0.756</b> (0.325)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.516</b> (0.065)	<b>0.720</b> (0.629)	<b>0.724</b> (0.633)
average of systems	<b>0.308</b> (0.140)	<b>0.476</b> (0.203)	<b>0.481</b> (0.206)

worst system	<b>0.026</b> (0.002)	<b>0.263</b> (0.017)	<b>0.263</b> (0.017)
best <u>baseline</u>	<b>0.556</b> (0.556)	<b>0.633</b> (0.633)	<b>0.643</b> (0.643)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.884</b> (0.881)	<b>0.930</b> (0.927)	<b>0.930</b> (0.927)
average of systems	<b>0.682</b> (0.395)	<b>0.733</b> (0.419)	<b>0.741</b> (0.423)
worst system	<b>0.382</b> (0.381)	<b>0.399</b> (0.397)	<b>0.407</b> (0.406)
best <u>baseline</u>	<b>0.756</b> (0.325)	<b>0.756</b> (0.325)	<b>0.756</b> (0.325)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.481</b> (0.060)	<b>0.501</b> (0.063)	<b>0.525</b> (0.066)
average of systems	<b>0.389</b> (0.067)	<b>0.426</b> (0.075)	<b>0.438</b> (0.076)
worst system	<b>0.296</b> (0.073)	<b>0.352</b> (0.087)	<b>0.352</b> (0.087)

Detailed results**nouns**

All test items in files with **-n** suffix.

number of items in task:	2756
--------------------------	------

	all senses	main senses only
average polysemy:	9.167	5.381

	<u><b>fine-grained</b></u>	<u><b>coarse-grained</b></u>
<u>average entropy:</u>	<b>1.740</b>	<b>1.167</b>

All systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
<u>human</u>	<b>0.973</b> (0.973)	<b>0.975</b> (0.975)	<b>0.977</b> (0.977)
best system	<b>0.850</b> (0.850)	<b>0.886</b> (0.886)	<b>0.918</b> (0.918)
average of systems	<b>0.594</b> (0.463)	<b>0.702</b> (0.529)	<b>0.749</b> (0.565)
worst system	<b>0.214</b> (0.174)	<b>0.384</b> (0.311)	<b>0.437</b> (0.354)
best <u>baseline</u>	<b>0.738</b> (0.569)	<b>0.815</b> (0.629)	<b>0.879</b> (0.679)

A systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
best system	<b>0.637</b> (0.581)	<b>0.773</b> (0.083)	<b>0.811</b> (0.087)
average of systems	<b>0.445</b> (0.298)	<b>0.608</b> (0.384)	<b>0.662</b> (0.422)
worst system	<b>0.214</b> (0.174)	<b>0.384</b> (0.311)	<b>0.437</b> (0.354)
best <u>baseline</u>	<b>0.628</b> (0.625)	<b>0.675</b> (0.672)	<b>0.731</b> (0.731)

S systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
best system	<b>0.850</b> (0.850)	<b>0.886</b> (0.886)	<b>0.918</b> (0.918)
average of systems	<b>0.712</b> (0.619)	<b>0.772</b> (0.673)	<b>0.817</b> (0.711)
worst system	<b>0.433</b> (0.424)	<b>0.588</b> (0.576)	<b>0.627</b> (0.614)

<u>best baseline</u>	<b>0.738</b> (0.569)	<b>0.815</b> (0.629)	<b>0.879</b> (0.679)
----------------------	----------------------	----------------------	----------------------

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.649</b> (0.388)	<b>0.733</b> (0.438)	<b>0.767</b> (0.459)
average of systems	<b>0.562</b> (0.264)	<b>0.699</b> (0.316)	<b>0.730</b> (0.331)
worst system	<b>0.476</b> (0.139)	<b>0.666</b> (0.194)	<b>0.694</b> (0.202)

Detailed results**all-nouns**

All test items in nouns, plus test items in files with **-p** suffix that were tagged with noun sense tags in the key.

number of items in task:	3792
--------------------------	------

	all senses	main senses only
average polysemy:	10.949	7.445

	fine-grained	coarse-grained
average <u>entropy</u> :	1.832	1.363

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.976</b> (0.976)	<b>0.978</b> (0.978)	<b>0.980</b> (0.980)

best system	<b>0.839</b> (0.839)	<b>0.872</b> (0.872)	<b>0.896</b> (0.896)
average of systems	<b>0.582</b> (0.427)	<b>0.671</b> (0.478)	<b>0.708</b> (0.504)
worst system	<b>0.235</b> (0.203)	<b>0.359</b> (0.309)	<b>0.395</b> (0.340)
best <u>baseline</u>	<b>0.746</b> (0.558)	<b>0.804</b> (0.602)	<b>0.852</b> (0.638)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.624</b> (0.570)	<b>0.689</b> (0.629)	<b>0.727</b> (0.664)
average of systems	<b>0.422</b> (0.267)	<b>0.554</b> (0.334)	<b>0.596</b> (0.362)
worst system	<b>0.235</b> (0.203)	<b>0.359</b> (0.309)	<b>0.395</b> (0.340)
best <u>baseline</u>	<b>0.564</b> (0.561)	<b>0.604</b> (0.601)	<b>0.642</b> (0.642)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.839</b> (0.839)	<b>0.872</b> (0.872)	<b>0.896</b> (0.896)
average of systems	<b>0.712</b> (0.584)	<b>0.764</b> (0.625)	<b>0.799</b> (0.653)
worst system	<b>0.433</b> (0.308)	<b>0.588</b> (0.419)	<b>0.627</b> (0.447)
best <u>baseline</u>	<b>0.746</b> (0.558)	<b>0.804</b> (0.602)	<b>0.852</b> (0.638)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.649</b> (0.282)	<b>0.733</b> (0.318)	<b>0.767</b> (0.333)
average of systems	<b>0.528</b> (0.204)	<b>0.638</b> (0.242)	<b>0.664</b> (0.252)
worst system	<b>0.408</b> (0.125)	<b>0.543</b> (0.166)	<b>0.562</b> (0.172)

Detailed results**verbs**

All test items in files with -v suffix.

number of items in task:	2501
--------------------------	------

	all senses	main senses only
average polysemy:	7.791	4.994

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.859	1.496

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.950</b> (0.947)	<b>0.955</b> (0.952)	<b>0.957</b> (0.954)
best system	<b>0.705</b> (0.697)	<b>0.741</b> (0.733)	<b>0.755</b> (0.747)
average of systems	<b>0.545</b> (0.500)	<b>0.583</b> (0.532)	<b>0.602</b> (0.549)
worst system	<b>0.155</b> (0.106)	<b>0.259</b> (0.178)	<b>0.272</b> (0.187)
best <u>baseline</u>	<b>0.701</b> (0.700)	<b>0.727</b> (0.725)	<b>0.746</b> (0.744)



A systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.469</b> (0.443)	<b>0.526</b> (0.497)	<b>0.537</b> (0.507)
average of systems	<b>0.390</b> (0.316)	<b>0.448</b> (0.361)	<b>0.465</b> (0.375)
worst system	<b>0.155</b> (0.106)	<b>0.259</b> (0.178)	<b>0.272</b> (0.187)
best <u><a href="#">baseline</a></u>	<b>0.547</b> (0.545)	<b>0.582</b> (0.579)	<b>0.592</b> (0.589)

S systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.705</b> (0.697)	<b>0.741</b> (0.733)	<b>0.755</b> (0.747)
average of systems	<b>0.622</b> (0.592)	<b>0.650</b> (0.618)	<b>0.671</b> (0.637)
worst system	<b>0.392</b> (0.392)	<b>0.414</b> (0.414)	<b>0.426</b> (0.425)
best <u><a href="#">baseline</a></u>	<b>0.701</b> (0.700)	<b>0.727</b> (0.725)	<b>0.746</b> (0.744)

Detailed results**all-verbs**

All test items in [verbs](#), plus test items in files with **-p** suffix that were tagged with verb sense tags in the key.

number of items in task:	2907
--------------------------	------

	all senses	main senses only
average polysemy:	10.821	7.733

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
average <u><a href="#">entropy</a></u> :	2.056	1.723

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.954</b> (0.951)	<b>0.958</b> (0.956)	<b>0.960</b> (0.957)
best system	<b>0.714</b> (0.707)	<b>0.748</b> (0.741)	<b>0.761</b> (0.754)
average of systems	<b>0.457</b> (0.419)	<b>0.486</b> (0.444)	<b>0.500</b> (0.458)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u><a href="#">baseline</a></u>	<b>0.676</b> (0.675)	<b>0.699</b> (0.697)	<b>0.717</b> (0.715)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.490</b> (0.460)	<b>0.543</b> (0.509)	<b>0.554</b> (0.519)
average of systems	<b>0.278</b> (0.226)	<b>0.316</b> (0.255)	<b>0.327</b> (0.264)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u><a href="#">baseline</a></u>	<b>0.541</b> (0.539)	<b>0.574</b> (0.572)	<b>0.583</b> (0.581)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.714</b> (0.707)	<b>0.748</b> (0.741)	<b>0.761</b> (0.754)
average of systems	<b>0.628</b> (0.597)	<b>0.653</b> (0.621)	<b>0.672</b> (0.639)
worst system	<b>0.431</b> (0.430)	<b>0.451</b> (0.450)	<b>0.460</b> (0.460)
best <u><a href="#">baseline</a></u>	<b>0.676</b> (0.675)	<b>0.699</b> (0.697)	<b>0.717</b> (0.715)

Detailed results**adjectives**

All test items in files with **-a** suffix.

number of items in task:	1406
--------------------------	------

	all senses	main senses only
average polysemy:	6.760	4.576

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.658	1.236

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.966</b> (0.965)	<b>0.972</b> (0.972)	<b>0.973</b> (0.973)
best system	<b>0.782</b> (0.756)	<b>0.797</b> (0.771)	<b>0.799</b> (0.772)
average of systems	<b>0.614</b> (0.543)	<b>0.652</b> (0.575)	<b>0.667</b> (0.587)
worst system	<b>0.181</b> (0.129)	<b>0.312</b> (0.223)	<b>0.323</b> (0.230)
best <u>baseline</u>	<b>0.718</b> (0.717)	<b>0.737</b> (0.735)	<b>0.740</b> (0.738)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.557</b> (0.532)	<b>0.590</b> (0.564)	<b>0.629</b> (0.601)
average of systems	<b>0.448</b> (0.368)	<b>0.503</b> (0.411)	<b>0.525</b> (0.429)
worst system	<b>0.181</b> (0.129)	<b>0.312</b> (0.223)	<b>0.323</b> (0.230)
best <u>baseline</u>	<b>0.681</b> (0.681)	<b>0.694</b> (0.694)	<b>0.709</b> (0.709)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.782</b> (0.756)	<b>0.797</b> (0.771)	<b>0.799</b> (0.772)
average of systems	<b>0.697</b> (0.631)	<b>0.726</b> (0.657)	<b>0.737</b> (0.666)
worst system	<b>0.589</b> (0.538)	<b>0.631</b> (0.576)	<b>0.634</b> (0.579)
best <u>baseline</u>	<b>0.718</b> (0.717)	<b>0.737</b> (0.735)	<b>0.740</b> (0.738)

Detailed results**all-adjectives**

All test items in adjectives, plus test items in files with **-p** suffix that were tagged with adjective sense tags in the key.

number of items in task:	1750
--------------------------	------

	all senses	main senses only
average polysemy:	8.430	5.868

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
average entropy:	1.867	1.490

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.957</b> (0.956)	<b>0.962</b> (0.962)	<b>0.963</b> (0.963)
best system	<b>0.755</b> (0.734)	<b>0.766</b> (0.746)	<b>0.768</b> (0.747)
average of systems	<b>0.599</b> (0.533)	<b>0.631</b> (0.561)	<b>0.644</b> (0.571)
worst system	<b>0.214</b> (0.163)	<b>0.320</b> (0.244)	<b>0.329</b> (0.251)
best <u><a href="#">baseline</a></u>	<b>0.688</b> (0.686)	<b>0.703</b> (0.701)	<b>0.705</b> (0.704)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.536</b> (0.505)	<b>0.563</b> (0.530)	<b>0.595</b> (0.560)
average of systems	<b>0.438</b> (0.357)	<b>0.486</b> (0.395)	<b>0.504</b> (0.410)
worst system	<b>0.214</b> (0.163)	<b>0.320</b> (0.244)	<b>0.329</b> (0.251)
best <u><a href="#">baseline</a></u>	<b>0.615</b> (0.615)	<b>0.627</b> (0.627)	<b>0.629</b> (0.629)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.755</b> (0.734)	<b>0.766</b> (0.746)	<b>0.768</b> (0.747)
average of systems	<b>0.680</b> (0.621)	<b>0.704</b> (0.643)	<b>0.713</b> (0.652)
worst system	<b>0.606</b> (0.563)	<b>0.639</b> (0.594)	<b>0.642</b> (0.597)
best <u><a href="#">baseline</a></u>	<b>0.688</b> (0.686)	<b>0.703</b> (0.701)	<b>0.705</b> (0.704)

Detailed results**indeterminates**

All test items in files with **-p** suffix; the part of speech of the word to be disambiguated has not been predetermined.

number of items in task:	1785
--------------------------	------

	all senses	main senses only
average polysemy:	18.692	15.199

	fine-grained	coarse-grained
average entropy:	2.468	2.284

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.970</b> (0.969)	<b>0.972</b> (0.970)	<b>0.973</b> (0.971)
best system	<b>0.775</b> (0.775)	<b>0.793</b> (0.793)	<b>0.797</b> (0.797)
average of systems	<b>0.536</b> (0.423)	<b>0.551</b> (0.435)	<b>0.553</b> (0.437)
worst system	<b>0.249</b> (0.051)	<b>0.260</b> (0.019)	<b>0.260</b> (0.019)
best <u>baseline</u>	<b>0.656</b> (0.531)	<b>0.658</b> (0.533)	<b>0.661</b> (0.535)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.572</b> (0.518)	<b>0.633</b> (0.573)	<b>0.635</b> (0.575)
average of systems	<b>0.352</b> (0.231)	<b>0.373</b> (0.249)	<b>0.375</b> (0.251)
worst system	<b>0.257</b> (0.246)	<b>0.260</b> (0.019)	<b>0.260</b> (0.019)
best <u>baseline</u>	<b>0.425</b> (0.424)	<b>0.444</b> (0.443)	<b>0.445</b> (0.444)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.775</b> (0.775)	<b>0.793</b> (0.793)	<b>0.797</b> (0.797)
average of systems	<b>0.694</b> (0.594)	<b>0.705</b> (0.604)	<b>0.707</b> (0.606)
worst system	<b>0.555</b> (0.555)	<b>0.567</b> (0.567)	<b>0.569</b> (0.569)
best <u>baseline</u>	<b>0.656</b> (0.531)	<b>0.658</b> (0.533)	<b>0.661</b> (0.535)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.249</b> (0.051)	<b>0.262</b> (0.054)	<b>0.262</b> (0.054)
average of systems	<b>0.249</b> (0.051)	<b>0.262</b> (0.054)	<b>0.262</b> (0.054)
worst system	<b>0.249</b> (0.051)	<b>0.262</b> (0.054)	<b>0.262</b> (0.054)

Detailed results**determinates**

All test items for ambiguous words with a predetermined part of speech; the complement of

indeterminates.

number of items in task:	6663
--------------------------	------

	all senses	main senses only
average polysemy:	8.143	5.066

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.768	1.305

#### All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.963</b> (0.962)	<b>0.967</b> (0.966)	<b>0.969</b> (0.967)
best system	<b>0.770</b> (0.769)	<b>0.802</b> (0.793)	<b>0.818</b> (0.818)
average of systems	<b>0.558</b> (0.388)	<b>0.656</b> (0.428)	<b>0.690</b> (0.449)
worst system	<b>0.187</b> (0.139)	<b>0.326</b> (0.242)	<b>0.357</b> (0.265)
best <u>baseline</u>	<b>0.720</b> (0.719)	<b>0.753</b> (0.752)	<b>0.779</b> (0.778)

#### A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.556</b> (0.519)	<b>0.773</b> (0.034)	<b>0.811</b> (0.036)
average of systems	<b>0.421</b> (0.232)	<b>0.573</b> (0.282)	<b>0.612</b> (0.303)
worst system	<b>0.187</b> (0.139)	<b>0.326</b> (0.242)	<b>0.357</b> (0.265)
best <u>baseline</u>	<b>0.584</b> (0.581)	<b>0.622</b> (0.619)	<b>0.641</b> (0.638)



S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.770</b> (0.769)	<b>0.802</b> (0.793)	<b>0.818</b> (0.818)
average of systems	<b>0.661</b> (0.552)	<b>0.711</b> (0.587)	<b>0.741</b> (0.610)
worst system	<b>0.433</b> (0.176)	<b>0.588</b> (0.238)	<b>0.626</b> (0.562)
best <u>baseline</u>	<b>0.720</b> (0.719)	<b>0.753</b> (0.752)	<b>0.779</b> (0.778)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.649</b> (0.160)	<b>0.733</b> (0.181)	<b>0.767</b> (0.190)
average of systems	<b>0.562</b> (0.108)	<b>0.699</b> (0.130)	<b>0.730</b> (0.137)
worst system	<b>0.476</b> (0.057)	<b>0.666</b> (0.080)	<b>0.694</b> (0.084)

Detailed results**trainable-nouns**

The intersection of nouns and trainable.

number of items in task:	2199
--------------------------	------

	all senses	main senses only
average polysemy:	10.067	5.675

	fine-grained	coarse-grained
average <u>entropy</u> :	1.892	1.271

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.976</b> (0.975)	<b>0.978</b> (0.978)	<b>0.980</b> (0.980)
best system	<b>0.833</b> (0.833)	<b>0.869</b> (0.869)	<b>0.909</b> (0.909)
average of systems	<b>0.594</b> (0.501)	<b>0.692</b> (0.565)	<b>0.748</b> (0.609)
worst system	<b>0.204</b> (0.160)	<b>0.373</b> (0.292)	<b>0.440</b> (0.345)
best <u><a href="#">baseline</a></u>	<b>0.751</b> (0.751)	<b>0.815</b> (0.788)	<b>0.879</b> (0.850)

## A systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.642</b> (0.583)	<b>0.753</b> (0.084)	<b>0.800</b> (0.089)
average of systems	<b>0.448</b> (0.298)	<b>0.594</b> (0.377)	<b>0.661</b> (0.425)
worst system	<b>0.204</b> (0.160)	<b>0.373</b> (0.292)	<b>0.440</b> (0.345)
best <u><a href="#">baseline</a></u>	<b>0.615</b> (0.612)	<b>0.659</b> (0.656)	<b>0.706</b> (0.706)

## S systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.833</b> (0.833)	<b>0.869</b> (0.869)	<b>0.909</b> (0.909)
average of systems	<b>0.713</b> (0.687)	<b>0.767</b> (0.739)	<b>0.819</b> (0.785)
worst system	<b>0.438</b> (0.434)	<b>0.564</b> (0.559)	<b>0.612</b> (0.606)
best <u><a href="#">baseline</a></u>	<b>0.751</b> (0.751)	<b>0.815</b> (0.788)	<b>0.879</b> (0.850)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.649</b> (0.486)	<b>0.733</b> (0.549)	<b>0.767</b> (0.575)
average of systems	<b>0.535</b> (0.304)	<b>0.678</b> (0.365)	<b>0.712</b> (0.383)
worst system	<b>0.420</b> (0.122)	<b>0.622</b> (0.181)	<b>0.657</b> (0.191)

Detailed results**all-trainable-nouns**

The intersection of all-nouns and trainable.

number of items in task:	2914
--------------------------	------

	<u>all senses</u>	<u>main senses only</u>
average polysemy:	11.963	8.000

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.898	1.406

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.978</b> (0.977)	<b>0.980</b> (0.980)	<b>0.982</b> (0.981)
best system	<b>0.829</b> (0.829)	<b>0.856</b> (0.856)	<b>0.887</b> (0.887)
average of systems	<b>0.588</b> (0.480)	<b>0.668</b> (0.529)	<b>0.713</b> (0.562)
worst system	<b>0.231</b> (0.193)	<b>0.353</b> (0.294)	<b>0.401</b> (0.335)

best baseline	<b>0.754</b> (0.754)	<b>0.804</b> (0.783)	<b>0.852</b> (0.830)
---------------	----------------------	----------------------	----------------------

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.687</b> (0.623)	<b>0.734</b> (0.666)	<b>0.783</b> (0.711)
average of systems	<b>0.426</b> (0.276)	<b>0.544</b> (0.337)	<b>0.597</b> (0.373)
worst system	<b>0.231</b> (0.193)	<b>0.353</b> (0.294)	<b>0.401</b> (0.335)
best baseline	<b>0.571</b> (0.568)	<b>0.607</b> (0.604)	<b>0.639</b> (0.636)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.829</b> (0.829)	<b>0.856</b> (0.856)	<b>0.887</b> (0.887)
average of systems	<b>0.723</b> (0.672)	<b>0.769</b> (0.713)	<b>0.810</b> (0.748)
worst system	<b>0.438</b> (0.327)	<b>0.564</b> (0.422)	<b>0.612</b> (0.457)
best baseline	<b>0.754</b> (0.754)	<b>0.804</b> (0.783)	<b>0.852</b> (0.830)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.649</b> (0.367)	<b>0.733</b> (0.414)	<b>0.767</b> (0.434)
average of systems	<b>0.507</b> (0.238)	<b>0.623</b> (0.284)	<b>0.653</b> (0.297)
worst system	<b>0.366</b> (0.110)	<b>0.513</b> (0.154)	<b>0.539</b> (0.161)

Detailed results

**untrainable-nouns**

The intersection of nouns and untrainable.

number of items in task:	557
--------------------------	-----

	all senses	main senses only
average polysemy:	5.616	4.219

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.140	0.755

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.964</b> (0.964)	<b>0.965</b> (0.965)	<b>0.966</b> (0.966)
best system	<b>0.916</b> (0.916)	<b>0.954</b> (0.954)	<b>0.955</b> (0.955)
average of systems	<b>0.543</b> (0.419)	<b>0.725</b> (0.524)	<b>0.727</b> (0.526)
worst system	<b>0.250</b> (0.230)	<b>0.422</b> (0.387)	<b>0.425</b> (0.390)
best <u>baseline</u>	<b>0.756</b> (0.756)	<b>0.828</b> (0.828)	<b>0.831</b> (0.831)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.617</b> (0.569)	<b>0.865</b> (0.081)	<b>0.865</b> (0.081)
average of systems	<b>0.428</b> (0.302)	<b>0.665</b> (0.410)	<b>0.666</b> (0.411)
worst system	<b>0.250</b> (0.230)	<b>0.422</b> (0.387)	<b>0.425</b> (0.390)

best <u>baseline</u>	<b>0.756</b> (0.756)	<b>0.828</b> (0.828)	<b>0.831</b> (0.831)
----------------------	----------------------	----------------------	----------------------

S systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
best system	<b>0.916</b> (0.916)	<b>0.954</b> (0.954)	<b>0.955</b> (0.955)
average of systems	<b>0.668</b> (0.601)	<b>0.787</b> (0.711)	<b>0.789</b> (0.713)
worst system	<b>0.415</b> (0.388)	<b>0.691</b> (0.645)	<b>0.692</b> (0.646)
best <u>baseline</u>	<b>0.681</b> (0.680)	<b>0.743</b> (0.741)	<b>0.746</b> (0.744)

O systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
best system	<b>0.691</b> (0.205)	<b>0.836</b> (0.248)	<b>0.836</b> (0.248)
average of systems	<b>0.691</b> (0.205)	<b>0.836</b> (0.248)	<b>0.836</b> (0.248)
worst system	<b>0.691</b> (0.205)	<b>0.836</b> (0.248)	<b>0.836</b> (0.248)

Detailed results**all-untrainable-nouns**

The intersection of all-nouns and untrainable.

number of items in task:	878
--------------------------	-----

	<u><b>all senses</b></u>	<u><b>main senses only</b></u>
average polysemy:	7.584	5.601

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
average <u><a href="#">entropy</a></u> :	1.614	1.217

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.972</b> (0.972)	<b>0.973</b> (0.973)	<b>0.973</b> (0.973)
best system	<b>0.871</b> (0.871)	<b>0.925</b> (0.925)	<b>0.926</b> (0.926)
average of systems	<b>0.496</b> (0.340)	<b>0.646</b> (0.419)	<b>0.647</b> (0.420)
worst system	<b>0.249</b> (0.236)	<b>0.376</b> (0.356)	<b>0.377</b> (0.358)
best <u><a href="#">baseline</a></u>	<b>0.603</b> (0.603)	<b>0.697</b> (0.697)	<b>0.699</b> (0.699)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.498</b> (0.326)	<b>0.742</b> (0.075)	<b>0.742</b> (0.075)
average of systems	<b>0.414</b> (0.238)	<b>0.599</b> (0.323)	<b>0.600</b> (0.323)
worst system	<b>0.249</b> (0.236)	<b>0.376</b> (0.356)	<b>0.377</b> (0.358)
best <u><a href="#">baseline</a></u>	<b>0.603</b> (0.603)	<b>0.697</b> (0.697)	<b>0.699</b> (0.699)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.871</b> (0.871)	<b>0.925</b> (0.925)	<b>0.926</b> (0.926)
average of systems	<b>0.596</b> (0.494)	<b>0.709</b> (0.573)	<b>0.710</b> (0.574)
worst system	<b>0.415</b> (0.246)	<b>0.585</b> (0.582)	<b>0.587</b> (0.584)
best <u><a href="#">baseline</a></u>	<b>0.535</b> (0.534)	<b>0.591</b> (0.589)	<b>0.592</b> (0.591)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.535</b> (0.176)	<b>0.633</b> (0.208)	<b>0.633</b> (0.208)
average of systems	<b>0.535</b> (0.176)	<b>0.633</b> (0.208)	<b>0.633</b> (0.208)
worst system	<b>0.535</b> (0.176)	<b>0.633</b> (0.208)	<b>0.633</b> (0.208)

Detailed results**trainable-adjectives**

The intersection of adjectives and trainable.

number of items in task:	1284
--------------------------	------

	all senses	main senses only
average polysemy:	6.927	4.536

	fine-grained	coarse-grained
average <u>entropy</u> :	1.700	1.237

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.964</b> (0.963)	<b>0.971</b> (0.970)	<b>0.972</b> (0.972)
best system	<b>0.766</b> (0.738)	<b>0.783</b> (0.754)	<b>0.785</b> (0.756)
average of systems	<b>0.603</b> (0.555)	<b>0.641</b> (0.588)	<b>0.657</b> (0.601)



worst system	<b>0.181</b> (0.142)	<b>0.312</b> (0.244)	<b>0.323</b> (0.252)
best <u>baseline</u>	<b>0.720</b> (0.719)	<b>0.740</b> (0.739)	<b>0.743</b> (0.743)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.549</b> (0.529)	<b>0.569</b> (0.548)	<b>0.612</b> (0.589)
average of systems	<b>0.424</b> (0.354)	<b>0.479</b> (0.398)	<b>0.503</b> (0.418)
worst system	<b>0.181</b> (0.142)	<b>0.312</b> (0.244)	<b>0.323</b> (0.252)
best <u>baseline</u>	<b>0.660</b> (0.660)	<b>0.677</b> (0.677)	<b>0.700</b> (0.700)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.766</b> (0.738)	<b>0.783</b> (0.754)	<b>0.785</b> (0.756)
average of systems	<b>0.692</b> (0.656)	<b>0.722</b> (0.683)	<b>0.734</b> (0.693)
worst system	<b>0.589</b> (0.589)	<b>0.631</b> (0.631)	<b>0.634</b> (0.634)
best <u>baseline</u>	<b>0.720</b> (0.719)	<b>0.740</b> (0.739)	<b>0.743</b> (0.743)

Detailed results**all-trainable-adjectives**

The intersection of all-adjectives and trainable.

number of items in task:	1628
--------------------------	------

	all senses	main senses only
average polysemy:	8.687	5.933

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
average <u><a href="#">entropy</a></u> :	1.915	1.510

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.955</b> (0.954)	<b>0.961</b> (0.960)	<b>0.962</b> (0.961)
best system	<b>0.740</b> (0.719)	<b>0.753</b> (0.731)	<b>0.755</b> (0.733)
average of systems	<b>0.589</b> (0.542)	<b>0.622</b> (0.570)	<b>0.635</b> (0.581)
worst system	<b>0.214</b> (0.176)	<b>0.320</b> (0.263)	<b>0.329</b> (0.270)
best <u><a href="#">baseline</a></u>	<b>0.686</b> (0.685)	<b>0.703</b> (0.702)	<b>0.706</b> (0.705)

[A systems](#)

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.529</b> (0.500)	<b>0.545</b> (0.515)	<b>0.579</b> (0.547)
average of systems	<b>0.418</b> (0.345)	<b>0.466</b> (0.384)	<b>0.485</b> (0.400)
worst system	<b>0.214</b> (0.176)	<b>0.320</b> (0.263)	<b>0.329</b> (0.270)
best <u><a href="#">baseline</a></u>	<b>0.594</b> (0.594)	<b>0.606</b> (0.606)	<b>0.609</b> (0.609)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.740</b> (0.719)	<b>0.753</b> (0.731)	<b>0.755</b> (0.733)
average of systems	<b>0.675</b> (0.640)	<b>0.700</b> (0.663)	<b>0.710</b> (0.672)
worst system	<b>0.606</b> (0.606)	<b>0.639</b> (0.639)	<b>0.642</b> (0.642)
best <u>baseline</u>	<b>0.686</b> (0.685)	<b>0.703</b> (0.702)	<b>0.706</b> (0.705)

Detailed results**all-untrainable-adjectives**

The intersection of all-adjectives and untrainable.

number of items in task:	122
--------------------------	-----

	all senses	main senses only
average polysemy:	5.000	5.000

	fine-grained	coarse-grained
average <u>entropy</u> :	1.224	1.224

**All systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.985</b> (0.985)	<b>0.985</b> (0.985)	<b>0.985</b> (0.985)
best system	<b>0.967</b> (0.967)	<b>0.967</b> (0.967)	<b>0.967</b> (0.967)
average of systems	<b>0.808</b> (0.699)	<b>0.838</b> (0.727)	<b>0.838</b> (0.727)

worst system	<b>0.617</b> (0.331)	<b>0.617</b> (0.331)	<b>0.617</b> (0.331)
best <u>baseline</u>	<b>0.902</b> (0.902)	<b>0.902</b> (0.902)	<b>0.902</b> (0.902)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.892</b> (0.877)	<b>0.892</b> (0.877)	<b>0.892</b> (0.877)
average of systems	<b>0.816</b> (0.646)	<b>0.862</b> (0.687)	<b>0.862</b> (0.687)
worst system	<b>0.639</b> (0.566)	<b>0.824</b> (0.730)	<b>0.824</b> (0.730)
best <u>baseline</u>	<b>0.902</b> (0.902)	<b>0.902</b> (0.902)	<b>0.902</b> (0.902)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.967</b> (0.967)	<b>0.967</b> (0.967)	<b>0.967</b> (0.967)
average of systems	<b>0.801</b> (0.742)	<b>0.819</b> (0.760)	<b>0.819</b> (0.760)
worst system	<b>0.617</b> (0.331)	<b>0.617</b> (0.331)	<b>0.617</b> (0.331)
best <u>baseline</u>	<b>0.704</b> (0.693)	<b>0.704</b> (0.693)	<b>0.704</b> (0.693)

Detailed results**trainable-indeterminates**

The intersection of indeterminates and trainable.

number of items in task:	1462
--------------------------	------

	all senses	main senses only
average polysemy:	20.392	16.789

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.475	2.343

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.967</b> (0.965)	<b>0.969</b> (0.967)	<b>0.969</b> (0.967)
best system	<b>0.776</b> (0.776)	<b>0.782</b> (0.782)	<b>0.784</b> (0.784)
average of systems	<b>0.538</b> (0.475)	<b>0.547</b> (0.482)	<b>0.549</b> (0.484)
worst system	<b>0.138</b> (0.009)	<b>0.138</b> (0.009)	<b>0.138</b> (0.009)
best <u>baseline</u>	<b>0.656</b> (0.649)	<b>0.658</b> (0.651)	<b>0.661</b> (0.654)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.685</b> (0.611)	<b>0.693</b> (0.619)	<b>0.696</b> (0.621)
average of systems	<b>0.335</b> (0.247)	<b>0.346</b> (0.256)	<b>0.348</b> (0.258)
worst system	<b>0.138</b> (0.009)	<b>0.138</b> (0.009)	<b>0.138</b> (0.009)
best <u>baseline</u>	<b>0.452</b> (0.451)	<b>0.466</b> (0.465)	<b>0.467</b> (0.466)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.776</b> (0.776)	<b>0.782</b> (0.782)	<b>0.784</b> (0.784)
average of systems	<b>0.713</b> (0.678)	<b>0.721</b> (0.686)	<b>0.723</b> (0.688)
worst system	<b>0.622</b> (0.622)	<b>0.633</b> (0.633)	<b>0.635</b> (0.635)
best <u>baseline</u>	<b>0.656</b> (0.649)	<b>0.658</b> (0.651)	<b>0.661</b> (0.654)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.210</b> (0.035)	<b>0.210</b> (0.035)	<b>0.210</b> (0.035)
average of systems	<b>0.210</b> (0.035)	<b>0.210</b> (0.035)	<b>0.210</b> (0.035)
worst system	<b>0.210</b> (0.035)	<b>0.210</b> (0.035)	<b>0.210</b> (0.035)

Detailed results**trainable-determinates**

The intersection of determinates and trainable.

number of items in task:	5984
--------------------------	------

	all senses	main senses only
average polysemy:	8.442	5.146

	fine-grained	coarse-grained
average <u>entropy</u> :	1.837	1.358

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.962</b> (0.961)	<b>0.967</b> (0.965)	<b>0.969</b> (0.967)
best system	<b>0.760</b> (0.751)	<b>0.790</b> (0.780)	<b>0.805</b> (0.795)
average of systems	<b>0.555</b> (0.398)	<b>0.644</b> (0.435)	<b>0.681</b> (0.458)
worst system	<b>0.180</b> (0.134)	<b>0.315</b> (0.234)	<b>0.349</b> (0.259)
best <u><a href="#">baseline</a></u>	<b>0.724</b> (0.723)	<b>0.755</b> (0.754)	<b>0.784</b> (0.783)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.549</b> (0.513)	<b>0.753</b> (0.031)	<b>0.800</b> (0.033)
average of systems	<b>0.420</b> (0.225)	<b>0.554</b> (0.270)	<b>0.600</b> (0.293)
worst system	<b>0.180</b> (0.134)	<b>0.315</b> (0.234)	<b>0.349</b> (0.259)
best <u><a href="#">baseline</a></u>	<b>0.573</b> (0.570)	<b>0.610</b> (0.607)	<b>0.631</b> (0.628)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.760</b> (0.751)	<b>0.790</b> (0.780)	<b>0.805</b> (0.795)
average of systems	<b>0.661</b> (0.576)	<b>0.705</b> (0.609)	<b>0.737</b> (0.635)
worst system	<b>0.438</b> (0.159)	<b>0.564</b> (0.205)	<b>0.612</b> (0.223)
best <u><a href="#">baseline</a></u>	<b>0.724</b> (0.723)	<b>0.755</b> (0.754)	<b>0.784</b> (0.783)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.649</b> (0.179)	<b>0.733</b> (0.202)	<b>0.767</b> (0.211)
average of systems	<b>0.535</b> (0.112)	<b>0.678</b> (0.134)	<b>0.712</b> (0.140)
worst system	<b>0.420</b> (0.045)	<b>0.622</b> (0.066)	<b>0.657</b> (0.070)

Detailed results**low-polysemy**

Items involving words whose polysemy is less than the median polysemy of 8.

number of items in task:	3771
--------------------------	------

	all senses	main senses only
average polysemy:	5.069	3.682

	fine-grained	coarse-grained
average <u>entropy</u> :	1.281	1.011

**All systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.971</b> (0.970)	<b>0.974</b> (0.973)	<b>0.975</b> (0.973)
best system	<b>0.839</b> (0.829)	<b>0.857</b> (0.846)	<b>0.865</b> (0.854)
average of systems	<b>0.613</b> (0.407)	<b>0.670</b> (0.431)	<b>0.683</b> (0.439)
worst system	<b>0.244</b> (0.178)	<b>0.368</b> (0.268)	<b>0.370</b> (0.270)



best <u>baseline</u>	<b>0.780</b> (0.678)	<b>0.798</b> (0.694)	<b>0.810</b> (0.704)
----------------------	----------------------	----------------------	----------------------

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.614</b> (0.583)	<b>0.661</b> (0.628)	<b>0.683</b> (0.649)
average of systems	<b>0.466</b> (0.248)	<b>0.553</b> (0.281)	<b>0.566</b> (0.289)
worst system	<b>0.244</b> (0.178)	<b>0.368</b> (0.268)	<b>0.370</b> (0.270)
best <u>baseline</u>	<b>0.675</b> (0.675)	<b>0.706</b> (0.706)	<b>0.731</b> (0.730)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.839</b> (0.829)	<b>0.857</b> (0.846)	<b>0.865</b> (0.854)
average of systems	<b>0.725</b> (0.580)	<b>0.756</b> (0.600)	<b>0.768</b> (0.610)
worst system	<b>0.522</b> (0.126)	<b>0.584</b> (0.513)	<b>0.590</b> (0.518)
best <u>baseline</u>	<b>0.780</b> (0.678)	<b>0.798</b> (0.694)	<b>0.810</b> (0.704)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.724</b> (0.106)	<b>0.782</b> (0.115)	<b>0.806</b> (0.118)
average of systems	<b>0.602</b> (0.080)	<b>0.681</b> (0.090)	<b>0.698</b> (0.093)
worst system	<b>0.481</b> (0.054)	<b>0.581</b> (0.066)	<b>0.590</b> (0.067)

Detailed results

**high-polysemy**

Items involving words whose polysemy is equal to or greater than the median polysemy of 8.

number of items in task:	4677
--------------------------	------

	all senses	main senses only
average polysemy:	14.648	10.050

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.428	1.916

**All systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.959</b> (0.957)	<b>0.963</b> (0.962)	<b>0.965</b> (0.964)
best system	<b>0.725</b> (0.725)	<b>0.760</b> (0.760)	<b>0.783</b> (0.783)
average of systems	<b>0.505</b> (0.352)	<b>0.604</b> (0.393)	<b>0.642</b> (0.416)
worst system	<b>0.178</b> (0.149)	<b>0.278</b> (0.233)	<b>0.316</b> (0.265)
best <u>baseline</u>	<b>0.624</b> (0.623)	<b>0.658</b> (0.657)	<b>0.689</b> (0.688)

**A systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.513</b> (0.467)	<b>0.686</b> (0.038)	<b>0.730</b> (0.040)
average of systems	<b>0.363</b> (0.201)	<b>0.512</b> (0.250)	<b>0.558</b> (0.273)
worst system	<b>0.178</b> (0.149)	<b>0.278</b> (0.233)	<b>0.316</b> (0.265)

best <u>baseline</u>	<b>0.465</b> (0.462)	<b>0.504</b> (0.501)	<b>0.527</b> (0.524)
----------------------	----------------------	----------------------	----------------------

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.725</b> (0.725)	<b>0.760</b> (0.760)	<b>0.783</b> (0.783)
average of systems	<b>0.615</b> (0.508)	<b>0.671</b> (0.545)	<b>0.705</b> (0.571)
worst system	<b>0.388</b> (0.148)	<b>0.572</b> (0.218)	<b>0.622</b> (0.238)
best <u>baseline</u>	<b>0.624</b> (0.623)	<b>0.658</b> (0.657)	<b>0.689</b> (0.688)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.611</b> (0.143)	<b>0.709</b> (0.166)	<b>0.748</b> (0.175)
average of systems	<b>0.486</b> (0.100)	<b>0.613</b> (0.124)	<b>0.645</b> (0.131)
worst system	<b>0.362</b> (0.057)	<b>0.517</b> (0.082)	<b>0.541</b> (0.086)

Detailed results**low-entropy**

Items involving words whose entropy is less than the median entropy of 1.85.

number of items in task:	3872
--------------------------	------

	all senses	main senses only
average polysemy:	7.309	5.472

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
average <u><a href="#">entropy</a></u> :	1.116	0.846

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.979</b> (0.978)	<b>0.981</b> (0.980)	<b>0.981</b> (0.981)
best system	<b>0.912</b> (0.912)	<b>0.929</b> (0.918)	<b>0.934</b> (0.923)
average of systems	<b>0.644</b> (0.452)	<b>0.727</b> (0.487)	<b>0.742</b> (0.496)
worst system	<b>0.267</b> (0.213)	<b>0.410</b> (0.327)	<b>0.418</b> (0.333)
best <u><a href="#">baseline</a></u>	<b>0.877</b> (0.713)	<b>0.892</b> (0.725)	<b>0.907</b> (0.738)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.728</b> (0.671)	<b>0.765</b> (0.705)	<b>0.787</b> (0.726)
average of systems	<b>0.471</b> (0.282)	<b>0.601</b> (0.333)	<b>0.618</b> (0.342)
worst system	<b>0.267</b> (0.213)	<b>0.410</b> (0.327)	<b>0.418</b> (0.333)
best <u><a href="#">baseline</a></u>	<b>0.696</b> (0.696)	<b>0.728</b> (0.728)	<b>0.749</b> (0.749)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.912</b> (0.912)	<b>0.929</b> (0.918)	<b>0.934</b> (0.923)
average of systems	<b>0.784</b> (0.635)	<b>0.826</b> (0.660)	<b>0.840</b> (0.669)
worst system	<b>0.483</b> (0.193)	<b>0.664</b> (0.548)	<b>0.669</b> (0.552)
best <u><a href="#">baseline</a></u>	<b>0.877</b> (0.713)	<b>0.892</b> (0.725)	<b>0.907</b> (0.738)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.728</b> (0.164)	<b>0.803</b> (0.181)	<b>0.822</b> (0.185)
average of systems	<b>0.587</b> (0.120)	<b>0.696</b> (0.141)	<b>0.707</b> (0.144)
worst system	<b>0.446</b> (0.077)	<b>0.589</b> (0.102)	<b>0.592</b> (0.103)

Detailed results**high-entropy**

Items involving words whose entropy is equal to or greater than the median entropy of 1.85.

number of items in task:	4576
--------------------------	------

	all senses	main senses only
average polysemy:	12.963	8.675

	fine-grained	coarse-grained
average entropy:	2.593	2.075

**All systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.953</b> (0.950)	<b>0.957</b> (0.955)	<b>0.960</b> (0.957)
best system	<b>0.653</b> (0.607)	<b>0.702</b> (0.020)	<b>0.752</b> (0.021)
average of systems	<b>0.471</b> (0.312)	<b>0.552</b> (0.344)	<b>0.596</b> (0.368)

worst system	<b>0.152</b> (0.119)	<b>0.233</b> (0.182)	<b>0.270</b> (0.211)
best <u>baseline</u>	<b>0.579</b> (0.578)	<b>0.615</b> (0.614)	<b>0.643</b> (0.642)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.450</b> (0.013)	<b>0.702</b> (0.020)	<b>0.752</b> (0.021)
average of systems	<b>0.358</b> (0.170)	<b>0.478</b> (0.205)	<b>0.532</b> (0.227)
worst system	<b>0.152</b> (0.119)	<b>0.233</b> (0.182)	<b>0.270</b> (0.211)
best <u>baseline</u>	<b>0.454</b> (0.452)	<b>0.495</b> (0.492)	<b>0.519</b> (0.516)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.653</b> (0.607)	<b>0.690</b> (0.690)	<b>0.718</b> (0.717)
average of systems	<b>0.559</b> (0.460)	<b>0.606</b> (0.493)	<b>0.641</b> (0.519)
worst system	<b>0.367</b> (0.093)	<b>0.473</b> (0.119)	<b>0.534</b> (0.135)
best <u>baseline</u>	<b>0.579</b> (0.578)	<b>0.615</b> (0.614)	<b>0.643</b> (0.642)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.561</b> (0.095)	<b>0.654</b> (0.111)	<b>0.706</b> (0.120)
average of systems	<b>0.456</b> (0.067)	<b>0.564</b> (0.082)	<b>0.610</b> (0.088)
worst system	<b>0.350</b> (0.038)	<b>0.474</b> (0.052)	<b>0.514</b> (0.056)

Detailed results

**accident-n**

number of items in task:	267
--------------------------	-----

	all senses	main senses only
polysemy:	8	2

	fine-grained	coarse-grained
entropy:	1.430	0.571

**All systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.987</b> (0.987)	<b>0.988</b> (0.988)	<b>0.991</b> (0.991)
best system	<b>0.944</b> (0.944)	<b>0.954</b> (0.954)	<b>0.987</b> (0.734)
average of systems	<b>0.658</b> (0.594)	<b>0.803</b> (0.689)	<b>0.878</b> (0.752)
worst system	<b>0.229</b> (0.064)	<b>0.489</b> (0.488)	<b>0.556</b> (0.554)
best <u>baseline</u>	<b>0.789</b> (0.783)	<b>0.828</b> (0.828)	<b>0.963</b> (0.963)

**A systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.856</b> (0.801)	<b>0.884</b> (0.828)	<b>0.972</b> (0.525)
average of systems	<b>0.494</b> (0.389)	<b>0.725</b> (0.533)	<b>0.809</b> (0.594)
worst system	<b>0.247</b> (0.029)	<b>0.489</b> (0.488)	<b>0.556</b> (0.554)
best <u>baseline</u>	<b>0.753</b> (0.753)	<b>0.789</b> (0.789)	<b>0.933</b> (0.933)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.944</b> (0.944)	<b>0.954</b> (0.954)	<b>0.987</b> (0.734)
average of systems	<b>0.819</b> (0.802)	<b>0.876</b> (0.857)	<b>0.950</b> (0.929)
worst system	<b>0.423</b> (0.423)	<b>0.720</b> (0.720)	<b>0.808</b> (0.808)
best <u>baseline</u>	<b>0.789</b> (0.783)	<b>0.828</b> (0.828)	<b>0.963</b> (0.963)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.644</b> (0.484)	<b>0.767</b> (0.576)	<b>0.823</b> (0.618)
average of systems	<b>0.437</b> (0.274)	<b>0.715</b> (0.381)	<b>0.752</b> (0.404)
worst system	<b>0.229</b> (0.064)	<b>0.662</b> (0.186)	<b>0.680</b> (0.191)

Detailed results**behaviour-n**

number of items in task:	279
--------------------------	-----

	all senses	main senses only
polysemy:	3	2

	fine-grained	coarse-grained
<u>entropy</u> :	0.390	0.295



## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.973</b> (0.973)	<b>0.973</b> (0.973)	<b>0.973</b> (0.973)
best system	<b>0.964</b> (0.964)	<b>0.964</b> (0.964)	<b>0.964</b> (0.964)
average of systems	<b>0.771</b> (0.676)	<b>0.909</b> (0.759)	<b>0.909</b> (0.759)
worst system	<b>0.263</b> (0.251)	<b>0.673</b> (0.642)	<b>0.673</b> (0.642)
best <u><a href="#">baseline</a></u>	<b>0.946</b> (0.946)	<b>0.961</b> (0.961)	<b>0.961</b> (0.961)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.924</b> (0.910)	<b>0.959</b> (0.928)	<b>0.959</b> (0.928)
average of systems	<b>0.585</b> (0.414)	<b>0.853</b> (0.576)	<b>0.853</b> (0.576)
worst system	<b>0.263</b> (0.251)	<b>0.673</b> (0.642)	<b>0.673</b> (0.642)
best <u><a href="#">baseline</a></u>	<b>0.946</b> (0.946)	<b>0.961</b> (0.961)	<b>0.961</b> (0.961)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.964</b> (0.964)	<b>0.964</b> (0.964)	<b>0.964</b> (0.964)
average of systems	<b>0.918</b> (0.889)	<b>0.945</b> (0.915)	<b>0.945</b> (0.915)
worst system	<b>0.575</b> (0.575)	<b>0.796</b> (0.796)	<b>0.796</b> (0.796)
best <u><a href="#">baseline</a></u>	<b>0.946</b> (0.946)	<b>0.961</b> (0.961)	<b>0.961</b> (0.961)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.777</b> (0.747)	<b>0.951</b> (0.276)	<b>0.951</b> (0.276)
average of systems	<b>0.626</b> (0.442)	<b>0.914</b> (0.559)	<b>0.914</b> (0.559)
worst system	<b>0.475</b> (0.138)	<b>0.877</b> (0.843)	<b>0.877</b> (0.843)

Detailed results**bet-n**

number of items in task:	274
--------------------------	-----

	all senses	main senses only
polysemy:	15	9

	fine-grained	coarse-grained
entropy:	3.200	2.563

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.970</b> (0.970)	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)
best system	<b>0.661</b> (0.661)	<b>0.807</b> (0.807)	<b>0.869</b> (0.869)
average of systems	<b>0.471</b> (0.376)	<b>0.548</b> (0.434)	<b>0.593</b> (0.474)
worst system	<b>0.132</b> (0.131)	<b>0.194</b> (0.135)	<b>0.194</b> (0.135)
best <u>baseline</u>	<b>0.547</b> (0.547)	<b>0.652</b> (0.621)	<b>0.782</b> (0.745)

A systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.629</b> (0.080)	<b>0.800</b> (0.102)	<b>0.800</b> (0.102)
average of systems	<b>0.394</b> (0.234)	<b>0.488</b> (0.288)	<b>0.533</b> (0.326)
worst system	<b>0.132</b> (0.131)	<b>0.194</b> (0.135)	<b>0.194</b> (0.135)
best <u><a href="#">baseline</a></u>	<b>0.406</b> (0.400)	<b>0.426</b> (0.404)	<b>0.452</b> (0.429)

S systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.661</b> (0.661)	<b>0.807</b> (0.807)	<b>0.869</b> (0.869)
average of systems	<b>0.515</b> (0.500)	<b>0.583</b> (0.565)	<b>0.635</b> (0.613)
worst system	<b>0.354</b> (0.354)	<b>0.413</b> (0.411)	<b>0.419</b> (0.418)
best <u><a href="#">baseline</a></u>	<b>0.547</b> (0.547)	<b>0.652</b> (0.621)	<b>0.782</b> (0.745)

O systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.615</b> (0.263)	<b>0.675</b> (0.288)	<b>0.675</b> (0.288)
average of systems	<b>0.550</b> (0.274)	<b>0.603</b> (0.299)	<b>0.616</b> (0.308)
worst system	<b>0.485</b> (0.285)	<b>0.531</b> (0.311)	<b>0.557</b> (0.327)

Detailed results**disability-n**

number of items in task:	160
--------------------------	-----

	all senses	main senses only
polysemy:	3	2

	fine-grained	coarse-grained
entropy:	1.053	0.457

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.959</b> (0.959)	<b>0.959</b> (0.959)	<b>0.959</b> (0.959)
best system	<b>0.900</b> (0.900)	<b>0.942</b> (0.906)	<b>0.942</b> (0.906)
average of systems	<b>0.735</b> (0.555)	<b>0.821</b> (0.625)	<b>0.821</b> (0.625)
worst system	<b>0.408</b> (0.400)	<b>0.484</b> (0.235)	<b>0.484</b> (0.235)
best <u>baseline</u>	<b>0.800</b> (0.800)	<b>0.938</b> (0.938)	<b>0.938</b> (0.938)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.870</b> (0.837)	<b>0.942</b> (0.906)	<b>0.942</b> (0.906)
average of systems	<b>0.718</b> (0.487)	<b>0.806</b> (0.556)	<b>0.806</b> (0.556)
worst system	<b>0.408</b> (0.400)	<b>0.484</b> (0.235)	<b>0.484</b> (0.235)
best <u>baseline</u>	<b>0.800</b> (0.800)	<b>0.938</b> (0.938)	<b>0.938</b> (0.938)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.900</b> (0.900)	<b>0.938</b> (0.938)	<b>0.938</b> (0.938)
average of systems	<b>0.745</b> (0.701)	<b>0.829</b> (0.782)	<b>0.829</b> (0.782)
worst system	<b>0.459</b> (0.450)	<b>0.567</b> (0.556)	<b>0.567</b> (0.556)
best <u>baseline</u>	<b>0.755</b> (0.750)	<b>0.906</b> (0.900)	<b>0.906</b> (0.900)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.815</b> (0.138)	<b>0.889</b> (0.150)	<b>0.889</b> (0.150)
average of systems	<b>0.815</b> (0.138)	<b>0.889</b> (0.150)	<b>0.889</b> (0.150)
worst system	<b>0.815</b> (0.138)	<b>0.889</b> (0.150)	<b>0.889</b> (0.150)

Detailed results**excess-n**

number of items in task:	186
--------------------------	-----

	all senses	main senses only
polysemy:	8	3

	fine-grained	coarse-grained
entropy:	2.387	1.199

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.973</b> (0.968)	<b>0.973</b> (0.968)	<b>0.973</b> (0.968)
best system	<b>0.882</b> (0.882)	<b>0.899</b> (0.899)	<b>0.919</b> (0.919)
average of systems	<b>0.495</b> (0.478)	<b>0.587</b> (0.551)	<b>0.692</b> (0.641)
worst system	<b>0.057</b> (0.013)	<b>0.236</b> (0.129)	<b>0.397</b> (0.216)
best <u>baseline</u>	<b>0.800</b> (0.796)	<b>0.820</b> (0.816)	<b>0.881</b> (0.876)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.456</b> (0.414)	<b>0.547</b> (0.497)	<b>0.822</b> (0.747)
average of systems	<b>0.267</b> (0.248)	<b>0.361</b> (0.337)	<b>0.530</b> (0.489)
worst system	<b>0.168</b> (0.167)	<b>0.236</b> (0.129)	<b>0.397</b> (0.216)
best <u>baseline</u>	<b>0.661</b> (0.661)	<b>0.747</b> (0.747)	<b>0.747</b> (0.747)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.882</b> (0.882)	<b>0.899</b> (0.899)	<b>0.919</b> (0.919)
average of systems	<b>0.662</b> (0.652)	<b>0.726</b> (0.712)	<b>0.801</b> (0.775)
worst system	<b>0.263</b> (0.161)	<b>0.396</b> (0.243)	<b>0.545</b> (0.545)
best <u>baseline</u>	<b>0.800</b> (0.796)	<b>0.820</b> (0.816)	<b>0.881</b> (0.876)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.532</b> (0.466)	<b>0.663</b> (0.580)	<b>0.723</b> (0.633)
average of systems	<b>0.294</b> (0.240)	<b>0.546</b> (0.337)	<b>0.609</b> (0.371)
worst system	<b>0.057</b> (0.013)	<b>0.429</b> (0.095)	<b>0.496</b> (0.109)

Detailed results**float-n**

number of items in task:	75
--------------------------	----

	all senses	main senses only
polysemy:	12	8

	fine-grained	coarse-grained
entropy:	2.340	2.042

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)	<b>0.993</b> (0.993)
best system	<b>0.813</b> (0.813)	<b>0.813</b> (0.813)	<b>0.840</b> (0.840)
average of systems	<b>0.367</b> (0.352)	<b>0.402</b> (0.379)	<b>0.493</b> (0.453)
worst system	<b>0.076</b> (0.067)	<b>0.076</b> (0.067)	<b>0.121</b> (0.107)
best <u>baseline</u>	<b>0.720</b> (0.720)	<b>0.720</b> (0.720)	<b>0.733</b> (0.733)

A systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.347</b> (0.347)	<b>0.347</b> (0.347)	<b>0.411</b> (0.189)
average of systems	<b>0.165</b> (0.145)	<b>0.200</b> (0.175)	<b>0.293</b> (0.254)
worst system	<b>0.076</b> (0.067)	<b>0.076</b> (0.067)	<b>0.121</b> (0.107)
best <u><a href="#">baseline</a></u>	<b>0.267</b> (0.260)	<b>0.281</b> (0.273)	<b>0.520</b> (0.520)

S systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.813</b> (0.813)	<b>0.813</b> (0.813)	<b>0.840</b> (0.840)
average of systems	<b>0.502</b> (0.498)	<b>0.534</b> (0.523)	<b>0.619</b> (0.597)
worst system	<b>0.107</b> (0.107)	<b>0.271</b> (0.271)	<b>0.357</b> (0.357)
best <u><a href="#">baseline</a></u>	<b>0.720</b> (0.720)	<b>0.720</b> (0.720)	<b>0.733</b> (0.733)

O systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.167</b> (0.053)	<b>0.236</b> (0.076)	<b>0.375</b> (0.120)
average of systems	<b>0.167</b> (0.053)	<b>0.236</b> (0.076)	<b>0.375</b> (0.120)
worst system	<b>0.167</b> (0.053)	<b>0.236</b> (0.076)	<b>0.375</b> (0.120)

Detailed results


---

**giant-n**



number of items in task:	118
--------------------------	-----

	all senses	main senses only
polysemy:	7	3

	fine-grained	coarse-grained
entropy:	2.054	1.239

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.975</b> (0.975)	<b>0.975</b> (0.975)	<b>0.992</b> (0.992)
best system	<b>0.856</b> (0.856)	<b>0.890</b> (0.890)	<b>0.983</b> (0.983)
average of systems	<b>0.487</b> (0.447)	<b>0.567</b> (0.512)	<b>0.710</b> (0.630)
worst system	<b>0.085</b> (0.085)	<b>0.085</b> (0.085)	<b>0.085</b> (0.085)
best <u>baseline</u>	<b>0.763</b> (0.763)	<b>0.839</b> (0.839)	<b>0.983</b> (0.983)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.492</b> (0.492)	<b>0.568</b> (0.568)	<b>0.720</b> (0.720)
average of systems	<b>0.271</b> (0.235)	<b>0.365</b> (0.307)	<b>0.525</b> (0.435)
worst system	<b>0.085</b> (0.085)	<b>0.085</b> (0.085)	<b>0.085</b> (0.085)
best <u>baseline</u>	<b>0.534</b> (0.534)	<b>0.571</b> (0.571)	<b>0.720</b> (0.720)

S systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.856</b> (0.856)	<b>0.890</b> (0.890)	<b>0.983</b> (0.983)
average of systems	<b>0.626</b> (0.605)	<b>0.701</b> (0.671)	<b>0.835</b> (0.790)
worst system	<b>0.274</b> (0.274)	<b>0.330</b> (0.330)	<b>0.460</b> (0.460)
best <u><a href="#">baseline</a></u>	<b>0.763</b> (0.763)	<b>0.839</b> (0.839)	<b>0.983</b> (0.983)

O systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.503</b> (0.414)	<b>0.557</b> (0.459)	<b>0.695</b> (0.572)
average of systems	<b>0.407</b> (0.245)	<b>0.465</b> (0.275)	<b>0.606</b> (0.350)
worst system	<b>0.310</b> (0.076)	<b>0.374</b> (0.092)	<b>0.517</b> (0.127)

Detailed results**knee-n**

number of items in task:	251
--------------------------	-----

	all senses	main senses only
polysemy:	22	12

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
entropy:	2.484	1.463

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.988</b> (0.988)	<b>0.988</b> (0.988)	<b>0.988</b> (0.988)
best system	<b>0.824</b> (0.821)	<b>0.958</b> (0.092)	<b>1.000</b> (0.096)
average of systems	<b>0.519</b> (0.466)	<b>0.708</b> (0.557)	<b>0.793</b> (0.626)
worst system	<b>0.000</b> (0.000)	<b>0.363</b> (0.131)	<b>0.462</b> (0.167)
best <u>baseline</u>	<b>0.665</b> (0.665)	<b>0.819</b> (0.803)	<b>0.861</b> (0.861)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.661</b> (0.582)	<b>0.958</b> (0.092)	<b>1.000</b> (0.096)
average of systems	<b>0.292</b> (0.243)	<b>0.623</b> (0.363)	<b>0.736</b> (0.447)
worst system	<b>0.000</b> (0.000)	<b>0.363</b> (0.131)	<b>0.462</b> (0.167)
best <u>baseline</u>	<b>0.578</b> (0.578)	<b>0.656</b> (0.656)	<b>0.833</b> (0.833)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.824</b> (0.821)	<b>0.877</b> (0.877)	<b>0.916</b> (0.916)
average of systems	<b>0.689</b> (0.667)	<b>0.765</b> (0.740)	<b>0.837</b> (0.806)
worst system	<b>0.394</b> (0.391)	<b>0.653</b> (0.648)	<b>0.756</b> (0.750)
best <u>baseline</u>	<b>0.665</b> (0.665)	<b>0.819</b> (0.803)	<b>0.861</b> (0.861)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.720</b> (0.447)	<b>0.768</b> (0.477)	<b>0.800</b> (0.497)
average of systems	<b>0.518</b> (0.266)	<b>0.747</b> (0.335)	<b>0.780</b> (0.350)
worst system	<b>0.316</b> (0.084)	<b>0.725</b> (0.193)	<b>0.761</b> (0.203)

Detailed results**onion-n**

number of items in task:	214
--------------------------	-----

	all senses	main senses only
polysemy:	4	4

	fine-grained	coarse-grained
entropy:	0.862	0.862

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.951</b> (0.951)	<b>0.951</b> (0.951)	<b>0.951</b> (0.951)
best system	<b>1.000</b> (0.019)	<b>1.000</b> (0.019)	<b>1.000</b> (0.019)
average of systems	<b>0.795</b> (0.639)	<b>0.795</b> (0.639)	<b>0.795</b> (0.639)
worst system	<b>0.434</b> (0.355)	<b>0.434</b> (0.355)	<b>0.434</b> (0.355)
best <u>baseline</u>	<b>0.911</b> (0.911)	<b>0.911</b> (0.911)	<b>0.911</b> (0.911)

A systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>1.000</b> (0.019)	<b>1.000</b> (0.019)	<b>1.000</b> (0.019)
average of systems	<b>0.710</b> (0.439)	<b>0.710</b> (0.439)	<b>0.710</b> (0.439)
worst system	<b>0.434</b> (0.355)	<b>0.434</b> (0.355)	<b>0.434</b> (0.355)
best <u><a href="#">baseline</a></u>	<b>0.911</b> (0.911)	<b>0.911</b> (0.911)	<b>0.911</b> (0.911)

S systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.925</b> (0.925)	<b>0.925</b> (0.925)	<b>0.925</b> (0.925)
average of systems	<b>0.857</b> (0.817)	<b>0.857</b> (0.817)	<b>0.857</b> (0.817)
worst system	<b>0.696</b> (0.696)	<b>0.696</b> (0.696)	<b>0.696</b> (0.696)
best <u><a href="#">baseline</a></u>	<b>0.909</b> (0.841)	<b>0.909</b> (0.841)	<b>0.909</b> (0.841)

O systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.858</b> (0.269)	<b>0.858</b> (0.269)	<b>0.858</b> (0.269)
average of systems	<b>0.810</b> (0.468)	<b>0.810</b> (0.468)	<b>0.810</b> (0.468)
worst system	<b>0.762</b> (0.668)	<b>0.762</b> (0.668)	<b>0.762</b> (0.668)

Detailed results**promise-n**

number of items in task:	113
--------------------------	-----

	all senses	main senses only
polysemy:	8	4

	fine-grained	coarse-grained
entropy:	1.851	0.963

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.965</b> (0.965)	<b>0.965</b> (0.965)	<b>0.965</b> (0.965)
best system	<b>0.867</b> (0.867)	<b>0.889</b> (0.889)	<b>0.920</b> (0.920)
average of systems	<b>0.566</b> (0.497)	<b>0.620</b> (0.541)	<b>0.705</b> (0.598)
worst system	<b>0.000</b> (0.000)	<b>0.124</b> (0.124)	<b>0.133</b> (0.133)
best <u>baseline</u>	<b>0.717</b> (0.717)	<b>0.746</b> (0.746)	<b>0.832</b> (0.832)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.630</b> (0.602)	<b>0.662</b> (0.633)	<b>0.759</b> (0.726)
average of systems	<b>0.413</b> (0.294)	<b>0.479</b> (0.338)	<b>0.588</b> (0.390)
worst system	<b>0.000</b> (0.000)	<b>0.124</b> (0.124)	<b>0.133</b> (0.133)
best <u>baseline</u>	<b>0.708</b> (0.708)	<b>0.737</b> (0.737)	<b>0.823</b> (0.823)

S systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.867</b> (0.867)	<b>0.889</b> (0.889)	<b>0.920</b> (0.920)
average of systems	<b>0.698</b> (0.677)	<b>0.737</b> (0.716)	<b>0.804</b> (0.779)
worst system	<b>0.362</b> (0.362)	<b>0.378</b> (0.378)	<b>0.419</b> (0.419)
best <u><a href="#">baseline</a></u>	<b>0.717</b> (0.717)	<b>0.746</b> (0.746)	<b>0.832</b> (0.832)

O systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.606</b> (0.538)	<b>0.748</b> (0.664)	<b>0.786</b> (0.698)
average of systems	<b>0.463</b> (0.336)	<b>0.555</b> (0.407)	<b>0.637</b> (0.451)
worst system	<b>0.319</b> (0.133)	<b>0.362</b> (0.150)	<b>0.489</b> (0.204)

Detailed results**rabbit-n**

number of items in task:	221
--------------------------	-----

	<u><a href="#">all senses</a></u>	<u><a href="#">main senses only</a></u>
<u><a href="#">polysemy</a></u> :	8	6

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
<u><a href="#">entropy</a></u> :	0.748	0.522

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.948</b> (0.948)	<b>0.950</b> (0.950)	<b>0.952</b> (0.952)
best system	<b>0.946</b> (0.946)	<b>0.955</b> (0.955)	<b>0.964</b> (0.964)
average of systems	<b>0.499</b> (0.417)	<b>0.809</b> (0.618)	<b>0.813</b> (0.621)
worst system	<b>0.000</b> (0.000)	<b>0.427</b> (0.425)	<b>0.427</b> (0.425)
best <u><a href="#">baseline</a></u>	<b>0.919</b> (0.919)	<b>0.928</b> (0.928)	<b>0.937</b> (0.937)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.921</b> (0.787)	<b>0.938</b> (0.722)	<b>0.941</b> (0.724)
average of systems	<b>0.385</b> (0.316)	<b>0.770</b> (0.531)	<b>0.773</b> (0.533)
worst system	<b>0.000</b> (0.000)	<b>0.427</b> (0.425)	<b>0.427</b> (0.425)
best <u><a href="#">baseline</a></u>	<b>0.919</b> (0.919)	<b>0.928</b> (0.928)	<b>0.937</b> (0.937)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.946</b> (0.946)	<b>0.955</b> (0.955)	<b>0.964</b> (0.964)
average of systems	<b>0.647</b> (0.588)	<b>0.846</b> (0.787)	<b>0.852</b> (0.792)
worst system	<b>0.372</b> (0.371)	<b>0.616</b> (0.209)	<b>0.619</b> (0.210)
best <u><a href="#">baseline</a></u>	<b>0.600</b> (0.600)	<b>0.624</b> (0.624)	<b>0.631</b> (0.631)



O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.500</b> (0.122)	<b>0.907</b> (0.222)	<b>0.907</b> (0.222)
average of systems	<b>0.500</b> (0.122)	<b>0.907</b> (0.222)	<b>0.907</b> (0.222)
worst system	<b>0.500</b> (0.122)	<b>0.907</b> (0.222)	<b>0.907</b> (0.222)

Detailed results**sack-n**

number of items in task:	82
--------------------------	----

	all senses	main senses only
polysemy:	11	9

	fine-grained	coarse-grained
entropy:	1.772	1.667

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)
best system	<b>0.878</b> (0.878)	<b>0.878</b> (0.878)	<b>0.878</b> (0.878)
average of systems	<b>0.594</b> (0.491)	<b>0.609</b> (0.505)	<b>0.609</b> (0.505)
worst system	<b>0.204</b> (0.067)	<b>0.204</b> (0.067)	<b>0.204</b> (0.067)
best <u>baseline</u>	<b>0.889</b> (0.878)	<b>0.889</b> (0.878)	<b>0.889</b> (0.878)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.812</b> (0.476)	<b>0.812</b> (0.476)	<b>0.812</b> (0.476)
average of systems	<b>0.522</b> (0.316)	<b>0.543</b> (0.336)	<b>0.543</b> (0.336)
worst system	<b>0.218</b> (0.207)	<b>0.328</b> (0.328)	<b>0.328</b> (0.328)
best <u>baseline</u>	<b>0.524</b> (0.524)	<b>0.524</b> (0.524)	<b>0.524</b> (0.524)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.878</b> (0.878)	<b>0.878</b> (0.878)	<b>0.878</b> (0.878)
average of systems	<b>0.672</b> (0.650)	<b>0.682</b> (0.659)	<b>0.682</b> (0.659)
worst system	<b>0.328</b> (0.328)	<b>0.339</b> (0.339)	<b>0.339</b> (0.339)
best <u>baseline</u>	<b>0.889</b> (0.878)	<b>0.889</b> (0.878)	<b>0.889</b> (0.878)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.693</b> (0.581)	<b>0.732</b> (0.614)	<b>0.732</b> (0.614)
average of systems	<b>0.449</b> (0.324)	<b>0.468</b> (0.340)	<b>0.468</b> (0.340)
worst system	<b>0.204</b> (0.067)	<b>0.204</b> (0.067)	<b>0.204</b> (0.067)

Detailed results


---

**scrap-n**

number of items in task:	156
--------------------------	-----

	all senses	main senses only
polysemy:	14	8

	fine-grained	coarse-grained
entropy:	2.839	1.999

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.965</b> (0.965)	<b>0.974</b> (0.974)	<b>0.978</b> (0.978)
best system	<b>0.686</b> (0.686)	<b>0.966</b> (0.179)	<b>0.966</b> (0.179)
average of systems	<b>0.490</b> (0.357)	<b>0.641</b> (0.456)	<b>0.696</b> (0.496)
worst system	<b>0.097</b> (0.096)	<b>0.161</b> (0.160)	<b>0.161</b> (0.160)
best <u>baseline</u>	<b>0.622</b> (0.622)	<b>0.760</b> (0.760)	<b>0.795</b> (0.795)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.586</b> (0.109)	<b>0.966</b> (0.179)	<b>0.966</b> (0.179)
average of systems	<b>0.474</b> (0.223)	<b>0.657</b> (0.297)	<b>0.718</b> (0.327)
worst system	<b>0.097</b> (0.096)	<b>0.161</b> (0.160)	<b>0.161</b> (0.160)
best <u>baseline</u>	<b>0.622</b> (0.622)	<b>0.760</b> (0.760)	<b>0.795</b> (0.795)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.686</b> (0.686)	<b>0.833</b> (0.833)	<b>0.865</b> (0.865)
average of systems	<b>0.525</b> (0.488)	<b>0.659</b> (0.611)	<b>0.713</b> (0.662)
worst system	<b>0.250</b> (0.250)	<b>0.256</b> (0.256)	<b>0.263</b> (0.263)
best <u>baseline</u>	<b>0.583</b> (0.583)	<b>0.708</b> (0.708)	<b>0.782</b> (0.782)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.489</b> (0.314)	<b>0.661</b> (0.423)	<b>0.698</b> (0.448)
average of systems	<b>0.352</b> (0.181)	<b>0.466</b> (0.242)	<b>0.499</b> (0.257)
worst system	<b>0.214</b> (0.048)	<b>0.271</b> (0.061)	<b>0.300</b> (0.067)

Detailed results**shirt-n**

number of items in task:	184
--------------------------	-----

	all senses	main senses only
polysemy:	8	6

	fine-grained	coarse-grained
entropy:	1.778	1.235

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.992</b> (0.992)	<b>0.995</b> (0.995)	<b>0.997</b> (0.997)
best system	<b>0.967</b> (0.967)	<b>0.978</b> (0.978)	<b>0.991</b> (0.592)
average of systems	<b>0.701</b> (0.554)	<b>0.782</b> (0.613)	<b>0.834</b> (0.653)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.858</b> (0.821)	<b>0.920</b> (0.880)	<b>0.983</b> (0.940)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.832</b> (0.337)	<b>0.940</b> (0.557)	<b>0.991</b> (0.592)
average of systems	<b>0.601</b> (0.376)	<b>0.712</b> (0.444)	<b>0.771</b> (0.481)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.821</b> (0.821)	<b>0.899</b> (0.899)	<b>0.940</b> (0.940)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.967</b> (0.967)	<b>0.978</b> (0.978)	<b>0.989</b> (0.989)
average of systems	<b>0.777</b> (0.723)	<b>0.840</b> (0.780)	<b>0.890</b> (0.826)
worst system	<b>0.441</b> (0.441)	<b>0.504</b> (0.504)	<b>0.555</b> (0.555)
best <u>baseline</u>	<b>0.858</b> (0.821)	<b>0.920</b> (0.880)	<b>0.983</b> (0.940)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.727</b> (0.574)	<b>0.806</b> (0.636)	<b>0.841</b> (0.664)
average of systems	<b>0.691</b> (0.338)	<b>0.748</b> (0.373)	<b>0.783</b> (0.389)
worst system	<b>0.655</b> (0.103)	<b>0.690</b> (0.109)	<b>0.724</b> (0.114)

Detailed results**steering-n**

number of items in task:	176
--------------------------	-----

	all senses	main senses only
polysemy:	5	4

	fine-grained	coarse-grained
entropy:	1.712	1.319

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.989</b> (0.989)	<b>0.989</b> (0.989)	<b>0.989</b> (0.989)
best system	<b>0.949</b> (0.949)	<b>0.972</b> (0.972)	<b>0.972</b> (0.972)
average of systems	<b>0.405</b> (0.318)	<b>0.428</b> (0.335)	<b>0.428</b> (0.335)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.716</b> (0.716)	<b>0.775</b> (0.775)	<b>0.775</b> (0.775)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.685</b> (0.293)	<b>0.746</b> (0.319)	<b>0.746</b> (0.319)
average of systems	<b>0.184</b> (0.129)	<b>0.203</b> (0.141)	<b>0.203</b> (0.141)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.716</b> (0.716)	<b>0.775</b> (0.775)	<b>0.775</b> (0.775)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.949</b> (0.949)	<b>0.972</b> (0.972)	<b>0.972</b> (0.972)
average of systems	<b>0.604</b> (0.527)	<b>0.635</b> (0.552)	<b>0.635</b> (0.552)
worst system	<b>0.043</b> (0.034)	<b>0.043</b> (0.034)	<b>0.043</b> (0.034)
best <u>baseline</u>	<b>0.716</b> (0.716)	<b>0.744</b> (0.744)	<b>0.744</b> (0.744)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.774</b> (0.369)	<b>0.774</b> (0.369)	<b>0.774</b> (0.369)
average of systems	<b>0.774</b> (0.369)	<b>0.774</b> (0.369)	<b>0.774</b> (0.369)
worst system	<b>0.774</b> (0.369)	<b>0.774</b> (0.369)	<b>0.774</b> (0.369)

Detailed results


---

**amaze-v**

number of items in task:	70
--------------------------	----

	all senses	main senses only
polysemy:	1	1

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	0.000	0.000

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)
best system	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)
average of systems	<b>0.945</b> (0.874)	<b>0.945</b> (0.874)	<b>0.945</b> (0.874)
worst system	<b>0.765</b> (0.743)	<b>0.765</b> (0.743)	<b>0.765</b> (0.743)
best <u>baseline</u>	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>1.000</b> (0.971)	<b>1.000</b> (0.971)	<b>1.000</b> (0.971)
average of systems	<b>0.923</b> (0.771)	<b>0.923</b> (0.771)	<b>0.923</b> (0.771)
worst system	<b>0.807</b> (0.357)	<b>0.807</b> (0.357)	<b>0.807</b> (0.357)
best <u>baseline</u>	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)



S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)
average of systems	<b>0.956</b> (0.925)	<b>0.956</b> (0.925)	<b>0.956</b> (0.925)
worst system	<b>0.765</b> (0.743)	<b>0.765</b> (0.743)	<b>0.765</b> (0.743)
best <u>baseline</u>	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)

Detailed results**bet-v**

number of items in task:	117
--------------------------	-----

	all senses	main senses only
polysemy:	9	4

	fine-grained	coarse-grained
entropy:	2.349	1.581

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.924</b> (0.916)	<b>0.932</b> (0.925)	<b>0.932</b> (0.925)
best system	<b>0.769</b> (0.769)	<b>0.778</b> (0.778)	<b>0.812</b> (0.812)
average of systems	<b>0.345</b> (0.336)	<b>0.400</b> (0.385)	<b>0.508</b> (0.481)
worst system	<b>0.009</b> (0.009)	<b>0.009</b> (0.009)	<b>0.009</b> (0.009)

best baseline	<b>0.714</b> (0.714)	<b>0.726</b> (0.726)	<b>0.803</b> (0.803)
---------------	----------------------	----------------------	----------------------

A systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
best system	<b>0.509</b> (0.479)	<b>0.518</b> (0.487)	<b>0.545</b> (0.513)
average of systems	<b>0.207</b> (0.190)	<b>0.261</b> (0.238)	<b>0.350</b> (0.317)
worst system	<b>0.034</b> (0.034)	<b>0.078</b> (0.077)	<b>0.129</b> (0.128)
best baseline	<b>0.714</b> (0.714)	<b>0.726</b> (0.726)	<b>0.803</b> (0.803)

S systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
best system	<b>0.769</b> (0.769)	<b>0.778</b> (0.778)	<b>0.812</b> (0.812)
average of systems	<b>0.414</b> (0.409)	<b>0.470</b> (0.458)	<b>0.587</b> (0.563)
worst system	<b>0.009</b> (0.009)	<b>0.009</b> (0.009)	<b>0.009</b> (0.009)
best baseline	<b>0.692</b> (0.692)	<b>0.701</b> (0.701)	<b>0.795</b> (0.795)

Detailed results**bother-v**

number of items in task:	209
--------------------------	-----

	<b>all senses</b>	<b>main senses only</b>
polysemy:	8	6

	<u><b>fine-grained</b></u>	<u><b>coarse-grained</b></u>
<u>entropy:</u>	<b>2.168</b>	<b>1.837</b>

All systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
<u>human</u>	<b>0.976</b> (0.976)	<b>0.976</b> (0.976)	<b>0.976</b> (0.976)
best system	<b>0.852</b> (0.852)	<b>0.871</b> (0.871)	<b>0.871</b> (0.871)
average of systems	<b>0.595</b> (0.567)	<b>0.623</b> (0.592)	<b>0.623</b> (0.592)
worst system	<b>0.142</b> (0.134)	<b>0.168</b> (0.158)	<b>0.168</b> (0.158)
best <u>baseline</u>	<b>0.632</b> (0.617)	<b>0.637</b> (0.622)	<b>0.637</b> (0.622)

A systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
best system	<b>0.491</b> (0.388)	<b>0.545</b> (0.431)	<b>0.545</b> (0.431)
average of systems	<b>0.379</b> (0.322)	<b>0.424</b> (0.361)	<b>0.424</b> (0.361)
worst system	<b>0.142</b> (0.134)	<b>0.168</b> (0.158)	<b>0.168</b> (0.158)
best <u>baseline</u>	<b>0.415</b> (0.405)	<b>0.467</b> (0.467)	<b>0.467</b> (0.467)

S systems

	<u><b>fine-grained</b></u> precision (recall)	<u><b>mixed-grained</b></u> precision (recall)	<u><b>coarse-grained</b></u> precision (recall)
best system	<b>0.852</b> (0.852)	<b>0.871</b> (0.871)	<b>0.871</b> (0.871)
average of systems	<b>0.703</b> (0.689)	<b>0.722</b> (0.708)	<b>0.722</b> (0.708)
worst system	<b>0.512</b> (0.512)	<b>0.541</b> (0.541)	<b>0.541</b> (0.541)
best <u>baseline</u>	<b>0.632</b> (0.617)	<b>0.637</b> (0.622)	<b>0.637</b> (0.622)

Detailed results**bury-v**

number of items in task:	201
--------------------------	-----

	all senses	main senses only
polysemy:	14	6

	fine-grained	coarse-grained
entropy:	2.759	2.401

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.928</b> (0.923)	<b>0.930</b> (0.925)	<b>0.933</b> (0.928)
best system	<b>0.562</b> (0.562)	<b>0.570</b> (0.570)	<b>0.587</b> (0.587)
average of systems	<b>0.396</b> (0.366)	<b>0.421</b> (0.388)	<b>0.436</b> (0.402)
worst system	<b>0.241</b> (0.224)	<b>0.270</b> (0.270)	<b>0.289</b> (0.289)
best <u>baseline</u>	<b>0.552</b> (0.552)	<b>0.557</b> (0.557)	<b>0.567</b> (0.567)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.300</b> (0.124)	<b>0.318</b> (0.296)	<b>0.331</b> (0.136)
average of systems	<b>0.265</b> (0.205)	<b>0.298</b> (0.234)	<b>0.312</b> (0.244)

worst system	<b>0.241</b> (0.224)	<b>0.270</b> (0.270)	<b>0.289</b> (0.289)
best <u>baseline</u>	<b>0.365</b> (0.365)	<b>0.383</b> (0.383)	<b>0.384</b> (0.384)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.562</b> (0.562)	<b>0.570</b> (0.570)	<b>0.587</b> (0.587)
average of systems	<b>0.449</b> (0.430)	<b>0.469</b> (0.449)	<b>0.486</b> (0.465)
worst system	<b>0.398</b> (0.398)	<b>0.417</b> (0.417)	<b>0.423</b> (0.423)
best <u>baseline</u>	<b>0.552</b> (0.552)	<b>0.557</b> (0.557)	<b>0.567</b> (0.567)

Detailed results**calculate-v**

number of items in task:	218
--------------------------	-----

	all senses	main senses only
polysemy:	5	3

	fine-grained	coarse-grained
entropy:	0.982	0.864

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.954</b> (0.950)	<b>0.959</b> (0.954)	<b>0.959</b> (0.954)
best system	<b>0.922</b> (0.922)	<b>0.922</b> (0.922)	<b>0.922</b> (0.922)
average of systems	<b>0.678</b> (0.635)	<b>0.694</b> (0.649)	<b>0.694</b> (0.649)
worst system	<b>0.272</b> (0.271)	<b>0.272</b> (0.271)	<b>0.272</b> (0.271)
best <u><a href="#">baseline</a></u>	<b>0.904</b> (0.904)	<b>0.904</b> (0.904)	<b>0.904</b> (0.904)

[A systems](#)

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.512</b> (0.509)	<b>0.614</b> (0.518)	<b>0.614</b> (0.518)
average of systems	<b>0.459</b> (0.382)	<b>0.503</b> (0.421)	<b>0.503</b> (0.421)
worst system	<b>0.333</b> (0.257)	<b>0.333</b> (0.257)	<b>0.333</b> (0.257)
best <u><a href="#">baseline</a></u>	<b>0.493</b> (0.493)	<b>0.493</b> (0.493)	<b>0.493</b> (0.493)

[S systems](#)

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.922</b> (0.922)	<b>0.922</b> (0.922)	<b>0.922</b> (0.922)
average of systems	<b>0.788</b> (0.761)	<b>0.790</b> (0.763)	<b>0.790</b> (0.763)
worst system	<b>0.272</b> (0.271)	<b>0.272</b> (0.271)	<b>0.272</b> (0.271)
best <u><a href="#">baseline</a></u>	<b>0.904</b> (0.904)	<b>0.904</b> (0.904)	<b>0.904</b> (0.904)

[Detailed results](#)

**consume-v**

number of items in task:	186
--------------------------	-----

	all senses	main senses only
polysemy:	6	4

	fine-grained	coarse-grained
entropy:	2.218	1.677

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.944</b> (0.939)	<b>0.955</b> (0.950)	<b>0.958</b> (0.953)
best system	<b>0.535</b> (0.532)	<b>0.589</b> (0.586)	<b>0.632</b> (0.629)
average of systems	<b>0.412</b> (0.378)	<b>0.489</b> (0.448)	<b>0.532</b> (0.486)
worst system	<b>0.189</b> (0.188)	<b>0.400</b> (0.398)	<b>0.416</b> (0.414)
best <u>baseline</u>	<b>0.546</b> (0.543)	<b>0.608</b> (0.605)	<b>0.654</b> (0.651)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.397</b> (0.392)	<b>0.454</b> (0.215)	<b>0.513</b> (0.243)
average of systems	<b>0.332</b> (0.274)	<b>0.437</b> (0.366)	<b>0.474</b> (0.396)
worst system	<b>0.189</b> (0.188)	<b>0.400</b> (0.398)	<b>0.416</b> (0.414)
best <u>baseline</u>	<b>0.416</b> (0.414)	<b>0.500</b> (0.497)	<b>0.535</b> (0.532)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.535</b> (0.532)	<b>0.589</b> (0.586)	<b>0.632</b> (0.629)
average of systems	<b>0.452</b> (0.430)	<b>0.515</b> (0.488)	<b>0.562</b> (0.531)
worst system	<b>0.373</b> (0.371)	<b>0.438</b> (0.435)	<b>0.489</b> (0.484)
best <u>baseline</u>	<b>0.546</b> (0.543)	<b>0.608</b> (0.605)	<b>0.654</b> (0.651)

Detailed results**derive-v**

number of items in task:	217
--------------------------	-----

	all senses	main senses only
polysemy:	6	4

	fine-grained	coarse-grained
entropy:	1.955	1.731

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.965</b> (0.961)	<b>0.965</b> (0.961)	<b>0.965</b> (0.961)
best system	<b>0.671</b> (0.668)	<b>0.671</b> (0.668)	<b>0.673</b> (0.673)
average of systems	<b>0.504</b> (0.467)	<b>0.523</b> (0.484)	<b>0.530</b> (0.490)



worst system	<b>0.161</b> (0.161)	<b>0.249</b> (0.249)	<b>0.258</b> (0.258)
best <u>baseline</u>	<b>0.588</b> (0.585)	<b>0.588</b> (0.585)	<b>0.588</b> (0.585)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.526</b> (0.475)	<b>0.526</b> (0.475)	<b>0.526</b> (0.475)
average of systems	<b>0.387</b> (0.320)	<b>0.423</b> (0.353)	<b>0.433</b> (0.361)
worst system	<b>0.161</b> (0.161)	<b>0.249</b> (0.249)	<b>0.258</b> (0.258)
best <u>baseline</u>	<b>0.530</b> (0.530)	<b>0.530</b> (0.530)	<b>0.530</b> (0.530)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.671</b> (0.668)	<b>0.671</b> (0.668)	<b>0.673</b> (0.673)
average of systems	<b>0.563</b> (0.541)	<b>0.573</b> (0.550)	<b>0.578</b> (0.555)
worst system	<b>0.486</b> (0.484)	<b>0.498</b> (0.495)	<b>0.509</b> (0.507)
best <u>baseline</u>	<b>0.588</b> (0.585)	<b>0.588</b> (0.585)	<b>0.588</b> (0.585)

Detailed results**float-v**

number of items in task:	229
--------------------------	-----

	all senses	main senses only
polysemy:	16	11

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
<u><a href="#">entropy</a></u> :	3.333	2.632

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.927</b> (0.923)	<b>0.938</b> (0.934)	<b>0.943</b> (0.939)
best system	<b>0.555</b> (0.555)	<b>0.633</b> (0.633)	<b>0.655</b> (0.655)
average of systems	<b>0.369</b> (0.339)	<b>0.419</b> (0.383)	<b>0.442</b> (0.404)
worst system	<b>0.021</b> (0.017)	<b>0.037</b> (0.031)	<b>0.042</b> (0.035)
best <u><a href="#">baseline</a></u>	<b>0.524</b> (0.524)	<b>0.579</b> (0.579)	<b>0.616</b> (0.616)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.394</b> (0.204)	<b>0.471</b> (0.244)	<b>0.507</b> (0.263)
average of systems	<b>0.276</b> (0.225)	<b>0.327</b> (0.267)	<b>0.356</b> (0.291)
worst system	<b>0.021</b> (0.017)	<b>0.037</b> (0.031)	<b>0.042</b> (0.035)
best <u><a href="#">baseline</a></u>	<b>0.403</b> (0.400)	<b>0.467</b> (0.463)	<b>0.502</b> (0.498)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.555</b> (0.555)	<b>0.633</b> (0.633)	<b>0.655</b> (0.655)
average of systems	<b>0.422</b> (0.402)	<b>0.471</b> (0.448)	<b>0.489</b> (0.466)
worst system	<b>0.279</b> (0.279)	<b>0.285</b> (0.285)	<b>0.288</b> (0.288)

best <u>baseline</u>	<b>0.524</b> (0.524)	<b>0.579</b> (0.579)	<b>0.616</b> (0.616)
----------------------	----------------------	----------------------	----------------------

Detailed results**invade-v**

number of items in task:	207
--------------------------	-----

	all senses	main senses only
polysemy:	6	3

	fine-grained	coarse-grained
<u>entropy</u> :	2.195	1.518

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.921</b> (0.912)	<b>0.922</b> (0.913)	<b>0.924</b> (0.915)
best system	<b>0.598</b> (0.219)	<b>0.645</b> (0.645)	<b>0.676</b> (0.676)
average of systems	<b>0.477</b> (0.422)	<b>0.559</b> (0.489)	<b>0.590</b> (0.518)
worst system	<b>0.119</b> (0.034)	<b>0.364</b> (0.104)	<b>0.373</b> (0.106)
best <u>baseline</u>	<b>0.570</b> (0.570)	<b>0.643</b> (0.643)	<b>0.686</b> (0.686)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.598</b> (0.219)	<b>0.622</b> (0.227)	<b>0.645</b> (0.236)
average of systems	<b>0.411</b> (0.303)	<b>0.528</b> (0.378)	<b>0.550</b> (0.395)
worst system	<b>0.119</b> (0.034)	<b>0.364</b> (0.104)	<b>0.373</b> (0.106)
best <u>baseline</u>	<b>0.420</b> (0.418)	<b>0.495</b> (0.493)	<b>0.517</b> (0.514)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.580</b> (0.580)	<b>0.645</b> (0.645)	<b>0.676</b> (0.676)
average of systems	<b>0.511</b> (0.482)	<b>0.575</b> (0.545)	<b>0.611</b> (0.579)
worst system	<b>0.454</b> (0.454)	<b>0.514</b> (0.514)	<b>0.553</b> (0.551)
best <u>baseline</u>	<b>0.570</b> (0.570)	<b>0.643</b> (0.643)	<b>0.686</b> (0.686)

Detailed results**promise-v**

number of items in task:	224
--------------------------	-----

	all senses	main senses only
polysemy:	6	3

	fine-grained	coarse-grained
entropy:	0.982	0.812

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.953</b> (0.953)	<b>0.962</b> (0.962)	<b>0.962</b> (0.962)
best system	<b>0.906</b> (0.906)	<b>0.913</b> (0.913)	<b>0.920</b> (0.920)
average of systems	<b>0.699</b> (0.655)	<b>0.745</b> (0.700)	<b>0.753</b> (0.708)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u><a href="#">baseline</a></u>	<b>0.862</b> (0.862)	<b>0.873</b> (0.873)	<b>0.884</b> (0.884)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.731</b> (0.728)	<b>0.772</b> (0.672)	<b>0.785</b> (0.683)
average of systems	<b>0.563</b> (0.483)	<b>0.681</b> (0.598)	<b>0.691</b> (0.607)
worst system	<b>0.058</b> (0.058)	<b>0.473</b> (0.473)	<b>0.478</b> (0.478)
best <u><a href="#">baseline</a></u>	<b>0.862</b> (0.862)	<b>0.873</b> (0.873)	<b>0.884</b> (0.884)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.906</b> (0.906)	<b>0.913</b> (0.913)	<b>0.920</b> (0.920)
average of systems	<b>0.766</b> (0.741)	<b>0.777</b> (0.751)	<b>0.785</b> (0.758)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u><a href="#">baseline</a></u>	<b>0.857</b> (0.857)	<b>0.868</b> (0.868)	<b>0.879</b> (0.879)

[Detailed results](#)

**sack-v.**

number of items in task:	178
--------------------------	-----

	all senses	main senses only
polysemy:	4	4

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	0.132	0.132

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.994</b> (0.994)	<b>0.994</b> (0.994)	<b>0.994</b> (0.994)
best system	<b>0.982</b> (0.646)	<b>0.982</b> (0.646)	<b>0.982</b> (0.646)
average of systems	<b>0.679</b> (0.646)	<b>0.679</b> (0.646)	<b>0.679</b> (0.646)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.907</b> (0.826)	<b>0.907</b> (0.826)	<b>0.907</b> (0.826)
average of systems	<b>0.301</b> (0.272)	<b>0.301</b> (0.272)	<b>0.301</b> (0.272)
worst system	<b>0.006</b> (0.006)	<b>0.006</b> (0.006)	<b>0.006</b> (0.006)
best <u>baseline</u>	<b>0.978</b> (0.978)	<b>0.978</b> (0.978)	<b>0.978</b> (0.978)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.982</b> (0.646)	<b>0.982</b> (0.646)	<b>0.982</b> (0.646)
average of systems	<b>0.868</b> (0.833)	<b>0.868</b> (0.834)	<b>0.868</b> (0.834)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)

Detailed results**scrap-v**

number of items in task:	186
--------------------------	-----

	all senses	main senses only
polysemy:	3	2

	fine-grained	coarse-grained
<u>entropy</u> :	0.694	0.133

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.981</b> (0.981)	<b>0.995</b> (0.995)	<b>0.995</b> (0.995)
best system	<b>0.898</b> (0.898)	<b>0.978</b> (0.978)	<b>0.978</b> (0.978)
average of systems	<b>0.693</b> (0.630)	<b>0.769</b> (0.699)	<b>0.769</b> (0.699)

worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.887</b> (0.887)	<b>0.978</b> (0.978)	<b>0.978</b> (0.978)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.862</b> (0.353)	<b>0.962</b> (0.395)	<b>0.962</b> (0.395)
average of systems	<b>0.607</b> (0.487)	<b>0.679</b> (0.545)	<b>0.679</b> (0.545)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.876</b> (0.876)	<b>0.978</b> (0.978)	<b>0.978</b> (0.978)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.898</b> (0.898)	<b>0.978</b> (0.978)	<b>0.978</b> (0.978)
average of systems	<b>0.735</b> (0.701)	<b>0.813</b> (0.776)	<b>0.813</b> (0.776)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.887</b> (0.887)	<b>0.978</b> (0.978)	<b>0.978</b> (0.978)

Detailed results**seize-v**

number of items in task:	259
--------------------------	-----

	all senses	main senses only
polysemy:	11	9



	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy:</u>	2.806	2.576

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.921</b> (0.921)	<b>0.921</b> (0.921)	<b>0.929</b> (0.929)
best system	<b>0.709</b> (0.641)	<b>0.709</b> (0.641)	<b>0.752</b> (0.680)
average of systems	<b>0.485</b> (0.440)	<b>0.485</b> (0.440)	<b>0.520</b> (0.471)
worst system	<b>0.123</b> (0.058)	<b>0.123</b> (0.058)	<b>0.197</b> (0.093)
best <u>baseline</u>	<b>0.680</b> (0.680)	<b>0.680</b> (0.680)	<b>0.710</b> (0.710)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.353</b> (0.129)	<b>0.353</b> (0.129)	<b>0.370</b> (0.135)
average of systems	<b>0.255</b> (0.192)	<b>0.255</b> (0.192)	<b>0.280</b> (0.207)
worst system	<b>0.123</b> (0.058)	<b>0.123</b> (0.058)	<b>0.197</b> (0.093)
best <u>baseline</u>	<b>0.498</b> (0.498)	<b>0.498</b> (0.498)	<b>0.498</b> (0.498)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.709</b> (0.641)	<b>0.709</b> (0.641)	<b>0.752</b> (0.680)
average of systems	<b>0.599</b> (0.564)	<b>0.599</b> (0.564)	<b>0.639</b> (0.603)
worst system	<b>0.402</b> (0.402)	<b>0.402</b> (0.402)	<b>0.483</b> (0.483)

best <u>baseline</u>	<b>0.680</b> (0.680)	<b>0.680</b> (0.680)	<b>0.710</b> (0.710)
----------------------	----------------------	----------------------	----------------------

Detailed results**brilliant-a**

number of items in task:	229
--------------------------	-----

	all senses	main senses only
polysemy:	10	8

	fine-grained	coarse-grained
entropy:	2.382	1.965

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.930</b> (0.926)	<b>0.954</b> (0.950)	<b>0.954</b> (0.950)
best system	<b>0.563</b> (0.563)	<b>0.642</b> (0.642)	<b>0.642</b> (0.642)
average of systems	<b>0.404</b> (0.370)	<b>0.474</b> (0.433)	<b>0.474</b> (0.433)
worst system	<b>0.062</b> (0.035)	<b>0.108</b> (0.061)	<b>0.108</b> (0.061)
best <u>baseline</u>	<b>0.512</b> (0.510)	<b>0.610</b> (0.608)	<b>0.610</b> (0.608)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.524</b> (0.485)	<b>0.599</b> (0.555)	<b>0.599</b> (0.555)
average of systems	<b>0.256</b> (0.224)	<b>0.307</b> (0.267)	<b>0.307</b> (0.267)
worst system	<b>0.062</b> (0.035)	<b>0.108</b> (0.061)	<b>0.108</b> (0.061)
best <u>baseline</u>	<b>0.476</b> (0.476)	<b>0.585</b> (0.585)	<b>0.585</b> (0.585)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.563</b> (0.563)	<b>0.642</b> (0.642)	<b>0.642</b> (0.642)
average of systems	<b>0.478</b> (0.443)	<b>0.557</b> (0.515)	<b>0.557</b> (0.515)
worst system	<b>0.388</b> (0.384)	<b>0.427</b> (0.424)	<b>0.427</b> (0.424)
best <u>baseline</u>	<b>0.512</b> (0.510)	<b>0.610</b> (0.608)	<b>0.610</b> (0.608)

Detailed results**deaf-a**

number of items in task:	122
--------------------------	-----

	all senses	main senses only
polysemy:	5	5

	fine-grained	coarse-grained
entropy:	1.224	1.224

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.985</b> (0.985)	<b>0.985</b> (0.985)	<b>0.985</b> (0.985)
best system	<b>0.967</b> (0.967)	<b>0.967</b> (0.967)	<b>0.967</b> (0.967)
average of systems	<b>0.808</b> (0.699)	<b>0.838</b> (0.727)	<b>0.838</b> (0.727)
worst system	<b>0.617</b> (0.331)	<b>0.617</b> (0.331)	<b>0.617</b> (0.331)
best <u><a href="#">baseline</a></u>	<b>0.902</b> (0.902)	<b>0.902</b> (0.902)	<b>0.902</b> (0.902)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.892</b> (0.877)	<b>0.892</b> (0.877)	<b>0.892</b> (0.877)
average of systems	<b>0.816</b> (0.646)	<b>0.862</b> (0.687)	<b>0.862</b> (0.687)
worst system	<b>0.639</b> (0.566)	<b>0.824</b> (0.730)	<b>0.824</b> (0.730)
best <u><a href="#">baseline</a></u>	<b>0.902</b> (0.902)	<b>0.902</b> (0.902)	<b>0.902</b> (0.902)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.967</b> (0.967)	<b>0.967</b> (0.967)	<b>0.967</b> (0.967)
average of systems	<b>0.801</b> (0.742)	<b>0.819</b> (0.760)	<b>0.819</b> (0.760)
worst system	<b>0.617</b> (0.331)	<b>0.617</b> (0.331)	<b>0.617</b> (0.331)
best <u><a href="#">baseline</a></u>	<b>0.704</b> (0.693)	<b>0.704</b> (0.693)	<b>0.704</b> (0.693)

[Detailed results](#)

**floating-a**

number of items in task:	47
--------------------------	----

	all senses	main senses only
polysemy:	5	4

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	1.748	1.539

**All systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.979</b> (0.979)	<b>0.979</b> (0.979)	<b>0.979</b> (0.979)
best system	<b>0.809</b> (0.809)	<b>0.809</b> (0.809)	<b>0.809</b> (0.809)
average of systems	<b>0.389</b> (0.373)	<b>0.405</b> (0.387)	<b>0.408</b> (0.389)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.660</b> (0.660)	<b>0.681</b> (0.681)	<b>0.681</b> (0.681)

**A systems**

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.578</b> (0.553)	<b>0.578</b> (0.553)	<b>0.578</b> (0.553)
average of systems	<b>0.216</b> (0.197)	<b>0.237</b> (0.214)	<b>0.241</b> (0.217)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.596</b> (0.596)	<b>0.596</b> (0.596)	<b>0.596</b> (0.596)

S systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.809</b> (0.809)	<b>0.809</b> (0.809)	<b>0.809</b> (0.809)
average of systems	<b>0.496</b> (0.484)	<b>0.510</b> (0.495)	<b>0.512</b> (0.497)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u><a href="#">baseline</a></u>	<b>0.660</b> (0.660)	<b>0.681</b> (0.681)	<b>0.681</b> (0.681)

Detailed results**generous-a**

number of items in task:	227
--------------------------	-----

	all senses	main senses only
polysemy:	6	6

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
<u><a href="#">entropy</a></u> :	2.303	2.303

All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.969</b> (0.969)	<b>0.969</b> (0.969)	<b>0.969</b> (0.969)
best system	<b>0.612</b> (0.612)	<b>0.612</b> (0.612)	<b>0.612</b> (0.612)
average of systems	<b>0.483</b> (0.434)	<b>0.483</b> (0.434)	<b>0.483</b> (0.434)

worst system	<b>0.294</b> (0.066)	<b>0.294</b> (0.066)	<b>0.294</b> (0.066)
best <u>baseline</u>	<b>0.488</b> (0.488)	<b>0.488</b> (0.488)	<b>0.488</b> (0.488)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.464</b> (0.432)	<b>0.464</b> (0.432)	<b>0.464</b> (0.432)
average of systems	<b>0.410</b> (0.307)	<b>0.410</b> (0.307)	<b>0.410</b> (0.307)
worst system	<b>0.294</b> (0.066)	<b>0.294</b> (0.066)	<b>0.294</b> (0.066)
best <u>baseline</u>	<b>0.407</b> (0.405)	<b>0.407</b> (0.405)	<b>0.407</b> (0.405)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.612</b> (0.612)	<b>0.612</b> (0.612)	<b>0.612</b> (0.612)
average of systems	<b>0.519</b> (0.498)	<b>0.519</b> (0.498)	<b>0.519</b> (0.498)
worst system	<b>0.383</b> (0.383)	<b>0.383</b> (0.383)	<b>0.383</b> (0.383)
best <u>baseline</u>	<b>0.488</b> (0.488)	<b>0.488</b> (0.488)	<b>0.488</b> (0.488)

Detailed results**giant-a**

number of items in task:	97
--------------------------	----

	all senses	main senses only
polysemy:	5	2

	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy:</u>	0.617	0.214

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)
best system	<b>0.990</b> (0.990)	<b>0.995</b> (0.995)	<b>1.000</b> (1.000)
average of systems	<b>0.643</b> (0.602)	<b>0.685</b> (0.644)	<b>0.688</b> (0.647)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.990</b> (0.990)	<b>0.995</b> (0.995)	<b>1.000</b> (1.000)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.571</b> (0.536)	<b>0.649</b> (0.649)	<b>0.649</b> (0.649)
average of systems	<b>0.195</b> (0.168)	<b>0.301</b> (0.273)	<b>0.301</b> (0.273)
worst system	<b>0.000</b> (0.000)	<b>0.011</b> (0.010)	<b>0.011</b> (0.010)
best <u>baseline</u>	<b>0.985</b> (0.985)	<b>0.995</b> (0.995)	<b>1.000</b> (1.000)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.990</b> (0.990)	<b>0.995</b> (0.995)	<b>1.000</b> (1.000)
average of systems	<b>0.867</b> (0.819)	<b>0.877</b> (0.830)	<b>0.881</b> (0.834)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)



best <u>baseline</u>	<b>0.990</b> (0.990)	<b>0.995</b> (0.995)	<b>1.000</b> (1.000)
----------------------	----------------------	----------------------	----------------------

Detailed results**modest-a**

number of items in task:	270
--------------------------	-----

	all senses	main senses only
polysemy:	9	3

	fine-grained	coarse-grained
entropy:	2.298	1.323

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.920</b> (0.920)	<b>0.935</b> (0.935)	<b>0.941</b> (0.941)
best system	<b>0.719</b> (0.719)	<b>0.728</b> (0.728)	<b>0.768</b> (0.477)
average of systems	<b>0.566</b> (0.529)	<b>0.608</b> (0.569)	<b>0.630</b> (0.588)
worst system	<b>0.086</b> (0.085)	<b>0.337</b> (0.309)	<b>0.368</b> (0.337)
best <u>baseline</u>	<b>0.648</b> (0.648)	<b>0.656</b> (0.656)	<b>0.670</b> (0.670)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.664</b> (0.644)	<b>0.669</b> (0.649)	<b>0.679</b> (0.659)
average of systems	<b>0.372</b> (0.315)	<b>0.449</b> (0.388)	<b>0.476</b> (0.411)
worst system	<b>0.086</b> (0.085)	<b>0.337</b> (0.309)	<b>0.368</b> (0.337)
best <u>baseline</u>	<b>0.637</b> (0.637)	<b>0.643</b> (0.643)	<b>0.656</b> (0.656)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.719</b> (0.719)	<b>0.728</b> (0.728)	<b>0.768</b> (0.477)
average of systems	<b>0.664</b> (0.637)	<b>0.687</b> (0.659)	<b>0.706</b> (0.676)
worst system	<b>0.567</b> (0.567)	<b>0.579</b> (0.579)	<b>0.596</b> (0.596)
best <u>baseline</u>	<b>0.648</b> (0.648)	<b>0.656</b> (0.656)	<b>0.670</b> (0.670)

Detailed results**slight-a**

number of items in task:	218
--------------------------	-----

	all senses	main senses only
polysemy:	6	3

	fine-grained	coarse-grained
entropy:	1.285	0.432

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.995</b> (0.995)	<b>0.995</b> (0.995)	<b>0.995</b> (0.995)
best system	<b>0.963</b> (0.963)	<b>0.963</b> (0.963)	<b>0.963</b> (0.963)
average of systems	<b>0.715</b> (0.672)	<b>0.782</b> (0.730)	<b>0.845</b> (0.784)
worst system	<b>0.028</b> (0.028)	<b>0.028</b> (0.028)	<b>0.028</b> (0.028)
best <u><a href="#">baseline</a></u>	<b>0.954</b> (0.954)	<b>0.954</b> (0.954)	<b>0.954</b> (0.954)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.595</b> (0.546)	<b>0.728</b> (0.674)	<b>0.845</b> (0.775)
average of systems	<b>0.468</b> (0.389)	<b>0.554</b> (0.459)	<b>0.660</b> (0.548)
worst system	<b>0.028</b> (0.028)	<b>0.028</b> (0.028)	<b>0.028</b> (0.028)
best <u><a href="#">baseline</a></u>	<b>0.954</b> (0.954)	<b>0.954</b> (0.954)	<b>0.954</b> (0.954)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.963</b> (0.963)	<b>0.963</b> (0.963)	<b>0.963</b> (0.963)
average of systems	<b>0.839</b> (0.814)	<b>0.896</b> (0.866)	<b>0.937</b> (0.903)
worst system	<b>0.682</b> (0.440)	<b>0.818</b> (0.528)	<b>0.908</b> (0.908)
best <u><a href="#">baseline</a></u>	<b>0.954</b> (0.954)	<b>0.954</b> (0.954)	<b>0.954</b> (0.954)

[Detailed results](#)

**wooden-a**

number of items in task:	196
--------------------------	-----

	all senses	main senses only
polysemy:	4	4

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
entropy:	0.365	0.365

## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)	<b>1.000</b> (1.000)
best system	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)
average of systems	<b>0.915</b> (0.849)	<b>0.915</b> (0.849)	<b>0.915</b> (0.849)
worst system	<b>0.515</b> (0.515)	<b>0.515</b> (0.515)	<b>0.515</b> (0.515)
best <u><a href="#">baseline</a></u>	<b>0.964</b> (0.964)	<b>0.964</b> (0.964)	<b>0.964</b> (0.964)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.958</b> (0.374)	<b>0.958</b> (0.374)	<b>0.958</b> (0.374)
average of systems	<b>0.830</b> (0.705)	<b>0.830</b> (0.705)	<b>0.830</b> (0.705)
worst system	<b>0.515</b> (0.515)	<b>0.515</b> (0.515)	<b>0.515</b> (0.515)
best <u><a href="#">baseline</a></u>	<b>0.949</b> (0.949)	<b>0.949</b> (0.949)	<b>0.949</b> (0.949)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)	<b>0.980</b> (0.980)
average of systems	<b>0.958</b> (0.921)	<b>0.958</b> (0.921)	<b>0.958</b> (0.921)
worst system	<b>0.938</b> (0.934)	<b>0.938</b> (0.934)	<b>0.938</b> (0.934)
best <u>baseline</u>	<b>0.964</b> (0.964)	<b>0.964</b> (0.964)	<b>0.964</b> (0.964)

Detailed results**band-p**

number of items in task:	302
--------------------------	-----

	all senses	main senses only
polysemy:	29	25

	fine-grained	coarse-grained
entropy:	1.749	1.669

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.990</b> (0.990)	<b>0.990</b> (0.990)	<b>0.990</b> (0.990)
best system	<b>0.904</b> (0.904)	<b>0.907</b> (0.907)	<b>0.907</b> (0.907)
average of systems	<b>0.616</b> (0.552)	<b>0.617</b> (0.552)	<b>0.617</b> (0.552)

worst system	<b>0.116</b> (0.116)	<b>0.116</b> (0.116)	<b>0.116</b> (0.116)
best <u>baseline</u>	<b>0.852</b> (0.843)	<b>0.852</b> (0.843)	<b>0.852</b> (0.843)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.883</b> (0.828)	<b>0.883</b> (0.828)	<b>0.883</b> (0.828)
average of systems	<b>0.294</b> (0.228)	<b>0.294</b> (0.228)	<b>0.295</b> (0.229)
worst system	<b>0.116</b> (0.116)	<b>0.116</b> (0.116)	<b>0.116</b> (0.116)
best <u>baseline</u>	<b>0.250</b> (0.248)	<b>0.257</b> (0.255)	<b>0.257</b> (0.255)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.904</b> (0.904)	<b>0.907</b> (0.907)	<b>0.907</b> (0.907)
average of systems	<b>0.864</b> (0.822)	<b>0.865</b> (0.822)	<b>0.865</b> (0.822)
worst system	<b>0.831</b> (0.765)	<b>0.831</b> (0.765)	<b>0.831</b> (0.765)
best <u>baseline</u>	<b>0.852</b> (0.843)	<b>0.852</b> (0.843)	<b>0.852</b> (0.843)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.388</b> (0.114)	<b>0.388</b> (0.114)	<b>0.388</b> (0.114)
average of systems	<b>0.388</b> (0.114)	<b>0.388</b> (0.114)	<b>0.388</b> (0.114)
worst system	<b>0.388</b> (0.114)	<b>0.388</b> (0.114)	<b>0.388</b> (0.114)

Detailed results

**bitter-p**

number of items in task:	373
--------------------------	-----

	all senses	main senses only
polysemy:	14	10

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
<u><a href="#">entropy</a></u> :	2.666	2.472

**All systems**

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.924</b> (0.917)	<b>0.927</b> (0.920)	<b>0.927</b> (0.920)
best system	<b>1.000</b> (0.008)	<b>1.000</b> (0.008)	<b>1.000</b> (0.008)
average of systems	<b>0.563</b> (0.456)	<b>0.578</b> (0.468)	<b>0.580</b> (0.470)
worst system	<b>0.240</b> (0.233)	<b>0.269</b> (0.261)	<b>0.273</b> (0.265)
best <u><a href="#">baseline</a></u>	<b>0.551</b> (0.550)	<b>0.556</b> (0.555)	<b>0.556</b> (0.555)

**A systems**

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.497</b> (0.440)	<b>0.526</b> (0.240)	<b>0.534</b> (0.244)
average of systems	<b>0.406</b> (0.323)	<b>0.434</b> (0.345)	<b>0.437</b> (0.347)
worst system	<b>0.240</b> (0.233)	<b>0.269</b> (0.261)	<b>0.273</b> (0.265)
best <u><a href="#">baseline</a></u>	<b>0.403</b> (0.402)	<b>0.414</b> (0.412)	<b>0.414</b> (0.412)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.729</b> (0.729)	<b>0.735</b> (0.735)	<b>0.737</b> (0.737)
average of systems	<b>0.597</b> (0.567)	<b>0.608</b> (0.576)	<b>0.610</b> (0.578)
worst system	<b>0.475</b> (0.475)	<b>0.477</b> (0.477)	<b>0.477</b> (0.477)
best <u>baseline</u>	<b>0.551</b> (0.550)	<b>0.556</b> (0.555)	<b>0.556</b> (0.555)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>1.000</b> (0.008)	<b>1.000</b> (0.008)	<b>1.000</b> (0.008)
average of systems	<b>1.000</b> (0.008)	<b>1.000</b> (0.008)	<b>1.000</b> (0.008)
worst system	<b>1.000</b> (0.008)	<b>1.000</b> (0.008)	<b>1.000</b> (0.008)

Detailed results**hurdle-p**

number of items in task:	323
--------------------------	-----

	all senses	main senses only
polysemy:	11	8

	fine-grained	coarse-grained
<u>entropy</u> :	2.437	2.019



## All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.985</b> (0.985)	<b>0.987</b> (0.987)	<b>0.987</b> (0.987)
best system	<b>0.793</b> (0.793)	<b>0.873</b> (0.873)	<b>0.873</b> (0.873)
average of systems	<b>0.405</b> (0.263)	<b>0.458</b> (0.308)	<b>0.458</b> (0.308)
worst system	<b>0.097</b> (0.093)	<b>0.269</b> (0.269)	<b>0.269</b> (0.269)
best <u><a href="#">baseline</a></u>	<b>0.334</b> (0.334)	<b>0.467</b> (0.467)	<b>0.467</b> (0.467)

[A](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.568</b> (0.065)	<b>0.568</b> (0.065)	<b>0.568</b> (0.065)
average of systems	<b>0.363</b> (0.162)	<b>0.428</b> (0.220)	<b>0.428</b> (0.220)
worst system	<b>0.097</b> (0.093)	<b>0.300</b> (0.300)	<b>0.300</b> (0.300)
best <u><a href="#">baseline</a></u>	<b>0.334</b> (0.334)	<b>0.467</b> (0.467)	<b>0.467</b> (0.467)

[S](#) systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.793</b> (0.793)	<b>0.873</b> (0.873)	<b>0.873</b> (0.873)
average of systems	<b>0.481</b> (0.432)	<b>0.520</b> (0.465)	<b>0.520</b> (0.465)
worst system	<b>0.251</b> (0.251)	<b>0.269</b> (0.269)	<b>0.269</b> (0.269)
best <u><a href="#">baseline</a></u>	<b>0.284</b> (0.283)	<b>0.328</b> (0.327)	<b>0.328</b> (0.327)

O systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
best system	<b>0.324</b> (0.125)	<b>0.360</b> (0.139)	<b>0.360</b> (0.139)
average of systems	<b>0.324</b> (0.125)	<b>0.360</b> (0.139)	<b>0.360</b> (0.139)
worst system	<b>0.324</b> (0.125)	<b>0.360</b> (0.139)	<b>0.360</b> (0.139)

Detailed results**sanction-p**

number of items in task:	431
--------------------------	-----

	<u><a href="#">all senses</a></u>	<u><a href="#">main senses only</a></u>
polysemy:	7	6

	<u><a href="#">fine-grained</a></u>	<u><a href="#">coarse-grained</a></u>
entropy:	1.810	1.722

All systems

	<u><a href="#">fine-grained</a></u> precision (recall)	<u><a href="#">mixed-grained</a></u> precision (recall)	<u><a href="#">coarse-grained</a></u> precision (recall)
<u><a href="#">human</a></u>	<b>0.981</b> (0.981)	<b>0.984</b> (0.984)	<b>0.984</b> (0.984)
best system	<b>0.865</b> (0.865)	<b>0.865</b> (0.865)	<b>0.865</b> (0.865)
average of systems	<b>0.548</b> (0.496)	<b>0.554</b> (0.502)	<b>0.554</b> (0.502)
worst system	<b>0.094</b> (0.030)	<b>0.094</b> (0.030)	<b>0.094</b> (0.030)
best <u><a href="#">baseline</a></u>	<b>0.781</b> (0.780)	<b>0.781</b> (0.780)	<b>0.781</b> (0.780)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.775</b> (0.703)	<b>0.775</b> (0.703)	<b>0.775</b> (0.703)
average of systems	<b>0.340</b> (0.268)	<b>0.346</b> (0.272)	<b>0.346</b> (0.272)
worst system	<b>0.145</b> (0.023)	<b>0.145</b> (0.023)	<b>0.145</b> (0.023)
best <u>baseline</u>	<b>0.585</b> (0.585)	<b>0.601</b> (0.601)	<b>0.601</b> (0.601)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.865</b> (0.865)	<b>0.865</b> (0.865)	<b>0.865</b> (0.865)
average of systems	<b>0.738</b> (0.703)	<b>0.746</b> (0.710)	<b>0.746</b> (0.710)
worst system	<b>0.592</b> (0.592)	<b>0.617</b> (0.617)	<b>0.617</b> (0.617)
best <u>baseline</u>	<b>0.781</b> (0.780)	<b>0.781</b> (0.780)	<b>0.781</b> (0.780)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.094</b> (0.030)	<b>0.094</b> (0.030)	<b>0.094</b> (0.030)
average of systems	<b>0.094</b> (0.030)	<b>0.094</b> (0.030)	<b>0.094</b> (0.030)
worst system	<b>0.094</b> (0.030)	<b>0.094</b> (0.030)	<b>0.094</b> (0.030)

Detailed results


---

**shake-p**

number of items in task:	356
--------------------------	-----

	all senses	main senses only
polysemy:	36	30

	fine-grained	coarse-grained
entropy:	3.696	3.531

## All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	<b>0.974</b> (0.974)	<b>0.977</b> (0.977)	<b>0.978</b> (0.978)
best system	<b>0.747</b> (0.747)	<b>0.767</b> (0.767)	<b>0.781</b> (0.781)
average of systems	<b>0.496</b> (0.456)	<b>0.510</b> (0.469)	<b>0.517</b> (0.476)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.632</b> (0.632)	<b>0.640</b> (0.640)	<b>0.649</b> (0.649)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.605</b> (0.328)	<b>0.621</b> (0.525)	<b>0.631</b> (0.534)
average of systems	<b>0.314</b> (0.254)	<b>0.334</b> (0.270)	<b>0.338</b> (0.274)
worst system	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)	<b>0.000</b> (0.000)
best <u>baseline</u>	<b>0.583</b> (0.581)	<b>0.609</b> (0.607)	<b>0.614</b> (0.612)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	<b>0.747</b> (0.747)	<b>0.767</b> (0.767)	<b>0.781</b> (0.781)
average of systems	<b>0.673</b> (0.643)	<b>0.684</b> (0.655)	<b>0.694</b> (0.665)
worst system	<b>0.559</b> (0.559)	<b>0.575</b> (0.575)	<b>0.587</b> (0.587)
best <u>baseline</u>	<b>0.632</b> (0.632)	<b>0.640</b> (0.640)	<b>0.649</b> (0.649)

Detailed results